EUREKA CLUSTER

ITEA 3 Software intensive systems and services

Update of the State-of-the-Art overview

Deliverable 1.3

POLicy Data Exploitation & Re-use

POLDER

PROJECT IDENTIFIER	POLDER
PROJECT TITLE	POLicy Data Exploitation & Re-use

DOCUMENT VERSION	1.0
PLANNED DELIVERY DATE	2020-12-31
ACTUAL DELIVERY DATE	2020-02-25
DOCUMENT TITLE	Update of the State-of-the-Art overview
WORK PACKAGE	WP1
DOCUMENT ACCESS	Public

ABSTRACT	The objective of the document is to provide a comprehensive state-of-the-art analysis of POLDER's targeted technologies and market. In addition to the up to date information regarding technology areas, the analysis will also include the related standards, ontologies and R&D projects. The document will also analyse most up-to-date needs of the cities.
KEYWORDS	

FUNCTION	NAME	ENTITY	
Editor	Yenal Ceylan	ForteArGe	
Editor	Serdar Sultanoglu	ForteArGe	
Editor	Migual Calero	TNP	
Editor	Güven Köse	Mantis	
Editor	Ali KAFALI	ACD	
Reviwer	Behzad Naderalvojoud	Mantis	

PARTNERS CONTRIBUTIONS RECORD

#	Partners	Contributor on Phase 1	Date of Contribution	Contributor on Phase 2	Date of Contribution2
1	ACCURO				
2	Acd				
3	ARD GROUP				
4	BEIA				
5	ForteArGe				
6	FCC				
7	Keypro Oy				
8	Mantis Software				
9	NOMMON				
10	Quobis				
11	Social And Beyond				
12	SIS				
13	Starflow				
14	TNP				
15	Vexiza				
16	Wellness Telecom				

CHANGES RECORD

VERSION	DATE	ENTITY	DESCRIPTION OF CHANGES
0.1	01.11.2020	ForteArGe	Compilation of available info & first inputs
0.2	08.12.2020	ForteArGe	Merge of information from cities needs analysis report from TNP
0.3	15.12.2020	ForteArGe	Updates regarding data visualisation, ontologies, standards and explainable AI.
0.4	17.12.2020	ACD	Updates sensors in IOT section 2.2.3
0.5	25.12.2020	Mantis	Adding the importance of social networks in smart city platforms.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. Introduction

1.1 Document Objectives And Scope

1.1.1 Objectives Of Document

The objective of the document is to provide a comprehensive state-of-the-art analysis of POLDER's targeted technologies and market. In addition to the up to date information regarding technology areas, the analysis will also include the related standards, ontologies and R&D projects. The document will also analyse most up-to-date needs of the cities.

1.2 Executive Summary

This document gives the state-of-the-art regarding the technologies related to POLDER project. It contains three main sections. The first section gives the current technologic and scientific situation in the project domain. The next section covers the standards related to POLDER technologies. The last section gives an in-depth analysis of the cities' requirements based on a survey conducted by University of Rotterdam.

2. State-Of-The-Art (SOTA) Analysis

2.1 Scientific State-Of-The-Art

The use of ICT in urban decision-making is already widely established. However, the incorporation of data, both technological (sensor) data and societal (citizen) data is still in its infancy. Poliakov and Van Veenstra (TNO) summarize the scientific state of the art; their paper is at the core of the POLDER project, so the following paragraphs are quotes from this.

Public sector innovation holds that "public policy and services need to become more open and innovative as well as being efficient and effective", making optimal use of ICTs [1]. As such, it encompasses a myriad of aspects. In [12] as much as fourteen aspects are identified of smartness in government, including evidence-based, technology savviness, openness, citizen engagement, and innovation. Public sector innovation means that public administrations operate as a platform [1,13,14] and use ICT to collaborate across organizational borders [15] and to involve citizens and other relevant stakeholders [6,7,16,17] with the purpose of creating public value [18] [19] [20]. Over the past decades, ICTs have had a great impact on services delivery [21], opened up public datasets [22] and increased citizens' participation [23]. The use of ICTs for policy making can, thus, be seen as a next step in public sector innovation [2].

The use of ICTs benefits policy making in two ways. The first is the use of new data sources, such as (real-time) sensor data, either physical (e.g. traffic monitoring [2] [4]), or virtual (e.g. social media data [2] [6]). "Data-driven decisions and intensive use of data, through ubiquitous sensing, advanced metering and integrated applications enable governments to make more informed decisions and improve the effectiveness of public policies and programs" [12]. Secondly, it requires from governments to collaborate across organizational borders and with citizens and businesses to

enable co-creation of policies [1,6,7,14]. “Co-creation is understood as the active flow and exchange of ideas, information, components and products across society (academia, government, business, civil society and citizens) which allows for a better understanding of participation, engagement and empowerment in policy development” [1].

Besides the deployment of ICTs to use new data sources and enable co-creation of policies, public sector innovation is concerned with the ability of public administrations to experiment and deploy ‘design-thinking’ in order to innovate [1,2,12]. Using innovative approaches such as gaming, simulation, and installing of sensors for do-it-yourself measurements, public administrations may experiment with policies [1,2]. In order to do, so many public administrations have set up Policy Labs [10,11]. “Policy Labs are emerging structures that construct public policies in an innovative, design-oriented fashion, in particular by engaging citizens and companies working within the public sector.” While Policy Labs exist in all shapes and sizes and on different levels of government (national, regional and municipal), the majority does not focus on a specific type of policy or on a specific phase of the policy cycle, but they employ a design and experimentation-based methodology of policy making [11].

Traditionally, open data has been provided by the public administration, as the opening of the administration data was considered to benefit business of the utilising companies, and thus the national economy, thus providing valuable information to the citizens. Recently, private companies have been interested in opening their own data, since open data has been identified as valuable, both in business and for a company’s internal use. [28]

2.2 Technological State-Of-The-Art

2.2.1 “SOTA” IN DATA

Smart cities are environments that are very data intensive. The term ‘Big data’ has been used for quite some time now; in general, it is a term for data that is so big that traditional processing with data becomes difficult. Typically, big data is characterized with five Vs, namely Volume, Velocity, Variety, Veracity and Value. The amount of data continues to rise with annual rise of 40%.

As the amount of data in increasing, need for practices to ensure quality of data increases. Relevant and irrelevant data regarding target environment is increasing constantly but capability to utilize information is insufficient.

Several ongoing activities are already targeted to creating future smart cities in Europe. In addition to strategic programs such as EU urban agenda digital transition activities, several member states have launched their own initiatives providing city data either at national level or via city cooperation (such as 6aika in Finland). The paradox of smart city is that although cities act as infrastructure and data hubs providing a variety of information to citizens and service providers, it is essentially companies that create the value-adding services to the smart city. Thus, city data should be designed to offer business value.

Although value of data in decision making is known there are relatively few studies in measuring business value of services based on the data. Schroder et al [24] have recently studied 28 industry

leaders and practitioners in order to identify key challenges, obstacles and policy suggestions in utilising data effectively. These include:

- **Data quality**: data quality is a major issue faced by big data users ([27]). Industry estimates are that between 75-90% of data scientist time is used to cleaning and structuring low quality data and perhaps 10% is spent doing actual data science. This problem can be partly mitigated if the development of and adherence to consistent standards is encouraged at both the organisation and industry level.
- **Context, metadata and data provenance**: "There is no easy or standard way to keep metadata about what the data means together with data in a nice and searchable and consistent way. And that means that the knowledge of what the data actually meant gets separated from the data". The problems created by this lack of standardisation are magnified because they limit the ability of skilled data workers to move seamlessly between industries. In order for data to remain useful in the future it is not only necessary that the data be readable, but also that it be documented in a transparent and consistent way so that all users understand what the data represents. A third point is that the provenance of third-party data is sometimes quite opaque.
- **Standards and accessibility**: It is not only in the application of metadata that standard practices are lacking. There is a more general lack of standardisation in the way that data is stored and processed. So long as a common standard that allows the interconnection of systems is absent, this will be a recurring challenge. A related issue is that of accessibility. Where tools do exist, they are often designed for implementation and use by specialist data scientists or engineers. More intensive use of data is only possible if tools are developed that allow non-expert workers to perform tasks that are currently the preserve of specialists.
- **Internal politics**: Company politics affect what data are shared internally, both between and within departments, as well as how data are shared with third parties. This very human element can create obstacles that technology alone cannot surmount. Decisions about how data are formatted for sharing and matching across data sets are critical and impact ease of later processing (c.f. the above discussions of data quality and metadata), but these decisions are often shaped as much by organisational structure and hierarchy as by practical or technical considerations.
- **The role of government**: There is a general desire for a minimum regulatory infrastructure combined with activities targeted at promoting the economic benefits of big data where these may not be well-known. There was agreement that big data policies should be transparent, clear, fair and consistent. These are hallmarks of any good regulation, but merit special mention because there is a shared sense that the existing regulatory environment fails on a number of these counts. One area of particular friction surrounds the issue of privacy and personal data.

To sum up, organizations experience challenges and failures in following issues, although these issues have been shown to provide large rewards when successful:

- Combining internal and external data for own operations and analytics; for instance, external data to outperform rivals with advanced data-driven operations (Walmart)
- Opening processes and data for external usage. For instance, opening processes and data to create an open supplier and consumer ecosystems (Apple)
- Selecting between physical and virtual implementation for optimal benefits. For instance, to speed-up integrations and enable operational analytics (Telefonica)
- Decoupling data and applications to increase agility and data quality. For instance, to enable advanced analytics without costly changes to processes or legacy systems (Dominos)
- Managing metadata to enable analytics while guaranteeing compliance. For instance, to support big data analytics in massive data lakes
- Achieving actual verifiable benefits from analytics investments. For instance, to enable A/B testing in business settings (DNA)
- Quality controlling internal and external data across processes. For instance, to guarantee regulatory compliance and validity of data-driven decision making.

Annex section provides detailed information about the state-of-the-art of data in smart tourism domain.

2.2.2 "SOTA" IN MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE

Like Big Data, the term Artificial Intelligence (AI) is a container term with unprecise definition. A key element in AI is Machine Learning which has recently accelerated, especially because Deep Learning Neural Networks have proven to be highly accurate and versatile. However, many issues need to be solved; understanding how a Neural Network actually works is an issue (explainability), and also the amount of training data needed to create a Neural Network can still be a costly exercise. Considered in the POLDER context, Deep Learning is a promising approach to process textual inputs, allowing the citizen involvement to be included in the modelling and simulation modules.

Other Machine Learning approaches exist, for instance Bayesian Belief networks, which allow explicit modelling of expert knowledge, and can then operate on a data set in order to adjust the model parameters (which is 'machine learning' at the core). This approach is very interesting to POLDER because it allows merging the pure 'data driven' approach with an 'expert-driven' approach.

Explainable AI

One aspect that is receiving a lot of attention recently, is Explainable AI (XAI). Several national initiatives exist ("VWData" in NL; "Transalgo" in FR), and the Big Data Value Association (bdva.eu) has adopted 'algorithmic transparency' as one of the core topics in 2018 to be addressed. Although transparency is mandated by the GDPR, it is not yet clear how detailed transparency should be pursued.

The concept 'explainability' can be elaborated into different directions, addressing different user groups:

- Designers: an algorithm developer (or model trainer) should be able to verify that the algorithm/model works as designed; therefore, the inner workings and detected relationships

should be transparent to the developer. Currently, some 'black box' approaches (notably Deep Learning) are still very hard to verify, even for developers. Approaches to handle this are now being developed, for instance by showing the 'attention' that a deep neural net pays to its inputs in specific situations.

- End users: end users should be able to understand how the outcome of a trained model relies on specific data inputs. It is important to note that this does not need to require specific data science knowledge. Approaches exist, for instance to explain when the output would change if one input parameter is adjusted; for instance, 'this person is denied to buy alcohol'; the output would change to 'allowed' if 'age' would be set to '18'. Another approach is to present a simplified model that is localized around the presented decision, for instance simple linear regression. (By avoiding the whole parameter space, which can be complex, the explanation is reduced to the current point of interest only.)
- Audience: explaining algorithms to the public at large has not yet been researched deeply; one problem is that it is hard to determine the a priori knowledge of the audience. Explanations can be done by showing what inputs have been used to reach a decision. To explain 'how' the decision has been made, approaches targeted at the end user need to be adopted.

POLDER will have to focus most on explainability towards end users and the larger audience, since that contributes to the goal of providing acceptable policy decisions. Research on the level of explainability (how efficient is an explanation) is nascent so POLDER should provide some pragmatic starting points here.

One of the practical starting point is to utilize post-hoc explainability techniques in the below table. These techniques are provided by [79], the work also provides comprehensive overview of explainable AI.

Text explanations	"deal with the problem of bringing explainability for a model by means of learning to generate text explanations that help explaining the results from the model. Text explanations also include every method generating symbols that represent the functioning of the model. These symbols may portrait the rationale of the algorithm by means of a semantic mapping from model to symbols"
Visual explanation	"for post-hoc explainability aim at visualizing the model's behavior. Many of the visualization methods existing in the literature come along with dimensionality reduction techniques that allow for a human interpretable simple visualization. Visualizations may be coupled with other techniques to improve their understanding, and are considered as the most suitable way to introduce complex interactions within the variables involved in the model to users not acquainted to ML modeling."
Local explanations	"tackle explainability by segmenting the solution space and giving explanations to less complex solution subspaces that are relevant for the whole model. These explanations can be formed by means of techniques with the differentiating property that these only explain part of the whole system's functioning."

Explanations by example	"consider the extraction of data examples that relate to the result generated by a certain model, enabling to get a better understanding of the model itself. Similarly to how humans behave when attempting to explain a given process, explanations by example are mainly centered in extracting representative examples that grasp the inner relationships and correlations found by the model being analyzed"
Explanations by simplification	"collectively denote those techniques in which a whole new system is rebuilt based on the trained model to be explained. This new, simplified model usually attempts at optimizing its resemblance to its antecedent functioning, while reducing its complexity, and keeping a similar performance score. An interesting by product of this family of post-hoc techniques is that the simplified model is, in general, easier to be implemented due to its reduced complexity with respect to the model it represents."
Feature relevance explanation	"for post-hoc explainability clarify the inner functioning of a model by computing a relevance score for its managed variables. These scores quantify the affection (sensitivity) a feature has upon the output of the model. A comparison of the scores among different variables unveils the importance granted by the model to each of such variables when producing its output. Feature relevance methods can be thought to be an indirect method to explain a model."

Table 1 Post-hoc explainability techniques by [79]

2.2.2.1 DESCRIPTION OF DESCRIPTIVE AND PREDICTIVE MODELS

In this sub-section the principles of the different models encountered for the state-of-the-art review of each use case are discussed.

2.2.2.1.1 Supervised learning

There is a great variety of models that allow different tourist indicators to be predicted, however, it is not possible to choose a type of model that surpasses the rest in terms of predictability. The results will depend on certain factors such as the type of data to be predicted, the number of dependent variables, the type of problem to be addressed, the type of industry and the techniques used.

The literature regarding the application of predictive models in tourism projects is reviewed below.

2.2.2.1.1.1 Temporal series

A time series is an ordered sequence of values of a random variable, consistently documented over time. These values allow the creation of stochastic models based on time series. This means that from the historical data of a variable, it is possible to create a base that allows to predict the future

values of the variable. Here are some examples of the use of this technique in the prediction of tourism indicators.

2.2.2.1.1.2 ARMA, ARIMA, SARIMA Models

The ARMA (autoregressive model of moving average), ARIMA (autoregressive integrated model of moving average) and SARIMA (autoregressive integrated model of seasonal moving average), are a type of stochastic models that include two models, one autoregressive and the other of moving average.

The ARMA model is applied when the data is stationary, while the ARIMA model considers the non-stationary component. The SARIMA model includes both the stationary and the seasonal components. In a simplified way, this type of model analyses the different components of the sample separately to predict each one separately. The series are disaggregated to obtain information on the trend, the seasonal component, and the noise of the series.

2.2.2.1.1.3 Regression Models

Regression models determine a predictive function from the calculation of a dependent variable from one or more independent or explanatory variables.

2.2.2.1.1.4 Support Vector Machines

Support vector machines or support vector machines (SVMs) are a set of supervised learning algorithms developed by Vladimir Vapnik and his team at AT&T laboratories.

More formally, an SVM constructs a hyperplane or set of hyperplanes in a very high (or even infinite) dimensional space that can be used in classification or regression problems. A good separation between the classes will allow a correct classification.

There are numerous studies in the field of tourism, which through the use of support vector machines have exceeded the prediction capacity of indicators such as the arrival of tourists.

2.2.2.1.1.5 Artificial Neuronal Networks

Artificial neural networks (also known as connectionist systems) are a computational model loosely inspired by observed behaviour in their biological counterpart. It consists of a set of units, called artificial neurons, connected together to transmit signals. The input information passes through the neural network (where it undergoes various operations) producing output values.

Each neuron is connected to others through links. In these links the output value of the previous neuron is multiplied by a weight value. These bond weights can increase or inhibit the activation state of adjacent neurons. Similarly, at the output of the neuron, there may be a limiting or threshold function, which modifies the resulting value or imposes a limit that must not be exceeded before propagating to another neuron. This function is known as the activation function.

These systems learn and train themselves, rather than being explicitly programmed, and excel in areas where finding solutions or features is difficult to express with conventional programming. To do this machine learning, you typically try to minimize a loss function that evaluates the network as a whole. The values of the weights of the neurons are updated seeking to reduce the value of the loss function. This process is done by propagating backwards.

2.2.2.1.2 Unsupervised learning

Clustering is a machine learning technique that involves grouping points. It is an unsupervised learning method and a famous technique for data analysis. It results in exhibiting similar properties at data points and different properties for different groups.

The goal of clustering is to extract value from large amounts of structured and unstructured data. It allows you to separate data based on their characteristics and group them into different groups according to their similarities. Clustering algorithms have a variety of uses in different industries.

Among the clustering techniques, the most typical are:

• K-means

• Mean Shift

• Expectation-Maximization (EM) Clustering using Gaussian Mixture Models (GMM)

• Hierarchical Clustering

2.2.2.2 Examples of applications to the use cases:

2.2.2.2.1 Temporal series

De Oliveira and Eduardo (2009) studied the prediction of tourist arrivals in Spain. To do this, they modelled tourism demand as a breakdown of components, according to variables such as country of residence, reason for the trip, type of transport and accommodation. They divided the arrival of tourists among 12 different countries of origin. They then added the different predictions of the

disaggregated components to obtain the total number of tourists. The result obtained through the disaggregated model was more accurate than that obtained through the use of traditional time series.

Abdelghany and Guzhva (2010) also used time series to analyze local passenger traffic at the Philadelphia airport. They analyzed the effect of seasonality, fuel prices, airline strategies, incidents, and financial factors to predict the level of activity at the airport. Its way of proceeding was to break down the total demand in local flights and connecting flights. This method offered better results in time periods close to the series (less than two years), although it can be used for longer periods (between two and five years), if certain assumptions are made.

2.2.2.2.2 ARMA, ARIMA, SARIMA Models

There are numerous studies that, through the use of this type of model, predicted tourist arrivals to different destinations, such as, for example, Andreoni and Postorino (2006) used the ARIMA model to forecast air demand based on historical data on volumes of traffic. passengers per year. In this case, the demand levels at the regional airport of Reggio Calabria in Italy were forecast with both time series models: univariate and multivariate.

The univariate model was generated from the time series of passenger volumes (both originated and destined) at Reggio Calabria airport from 1989 to 2003 and was later validated with data after 2003 that were not used in the model. For the multivariate model, the explanatory variables were the volume of passengers, as well as the distribution of the income level of passengers.

Both models offered results that adjusted to reality, obtaining more adjusted results for the univariate model. However, this does not mean that the univariate model is better. The predictability of the univariate model is highly dependent on the stability of the boundary conditions. This problem can be solved with the use of explanatory variables; however, it is not always possible to find time series of these explanatory variables.

2.2.2.2.3 Regression Models

Wu et al. (2012) studied the prediction of tourist demand in Hong Kong using regression models, based on the disaggregated data of 13 nationalities of tourist arrivals to Hong Kong, income data from the country of origin, price data from the country of origin in relation to with prices at destination, cost of transport between origin and destination, exchange rate between currencies, population of the issuing country, spending on marketing promotion of tourism from the issuing

country and a quality factor of the visit that takes into account possible events or policy changes regarding tourism. The scatter diagrams of the proposed model allowed a reduction in computational complexity, in addition to showing more effective results compared to other models such as ARIMA. On the other hand, Blainey and Mulley (2013) predicted railway demand in New South Wales, Australia. They tested numerous independent variables to generate the model, such as recruitment population, employment, income and age profile, household size, car, etc., to produce the best fit prediction model.

2.2.2.2.4 Support Vector Machines

For example, and Lin and Lee (2013) studied the prediction of tourist demand to Taiwan using different predictive models, obtaining a lower mean square error for the SVM model. For this, the dependent variable was the number of tourists, and the explanatory variables used were the population and income level of the countries of origin, the travel expenses of the tourists, the accommodation rate each night and the price level at the destination and the expenses of traveling to the tourist destination.

2.2.2.2.5 Artificial Neuronal Networks

Since the creation of this type of model for the prediction of tourist demand around 1990, the number of publications using this type of model has been growing over the last decade. For example, Fernandes and Teixeira (2008) studied the tourist demand in northern Portugal using a model that used the monthly values of tourist overnight stays, segmented by region of overnight stay, to calculate future values of monthly overnight stays by tourists. This type of model has been used to estimate the demand for cruises for the Izmir port in Turkey (Cuhadar et al., 2014), or air demand for domestic flights in Australia (Srisaeng et al., 2015a), among others.

2.2.3 "SOTA" IN IOT AND SENSORS

The Internet of Things (IoT) is understood as a network of internet-connected objects able to collect and exchange data using embedded sensors. These IoT devices can be connected with many entities, from remote devices, dashboards, data storage and analytics, networks and security, among others. They are able to enable systems, businesses, consumers and governments for monitoring, control, and decision-making purposes.

Current solutions in the field of smart cities use using sensors, may integrate information from public data sources and may work on-line or real-time. Typically, the most successful solutions are custom

built to solve a certain problem really well. However, they are not scalable, do not give the opportunity to integrate new data easily or are not wireless / real-time.

Smart energy, smart health, environmental management, smart transport, smart homes, community safety, condition monitoring and maintenance of infrastructure are the examples of smart city applications. IoT applications for these diverse domains were determined in the literature. In addition to this, a prototype technical solution for energy control and comfort in a home for proof of concept of a smart city infrastructure application was improved in a laboratory environment [80].

Big cities have some problems about emergency situations management which is a very actual topic. These smart cities should provide to citizens protection. Therefore, different types of sensor networks are created and implemented by researches to estimate, prevent and act fast and effectively on dangerous situations for monitoring applications. Because of these, various configurations of smart sensors networks were implemented and optimized in the studies and research. They can store and collect data from various sensor from an extended area like a city, in accordance with "IoT Smart city" concept [81].

Several cities have performed interesting experiences with IoT. A good example is Amsterdam, the offered solar panels and home energy storage units for households that are connected to the smart grid of Amsterdam. Residents can store energy during off-peak hours and even sell spare energy back to the grid. In Santander, more than 12,000 sensors are embedded to measure different things like the amount of trash in containers, the number of parking spaces available, or the size of crowds. In Copenhagen the bike traffic is being monitored using sensors. San Diego is monitoring pedestrian traffic with cameras in the streetlights and to reroute cars during peak hours, avoiding accidents with pedestrian and handling congestions. By installing a sensor network on air quality, sound, and microclimate, San Jose implemented an IoT Smart City Demonstration Platform for their Green Vision project. Wuhan has developed a sensor-powered traffic control system where automatic sensors and authorities (traffic police) are integrated together to optimize the traffic lights across the city.

There is a huge diversity of sensors useful for Smart Cities, according to the application areas. The general challenge in almost all cases is that cost and precision need to be balanced. This is the case for instance for environmental monitoring, (indoor) localization, imaging, and crowd tracking. The POLDER consortium contains sensor developers and vendors that are leading the developments in this area.

Today, the increase in the amount of energy used by the occupant in buildings increases every year. Therefore, a new Heating, ventilation, and air conditioning (HVAC) system is developed to control and monitor the energy consumption of buildings with fuzzy logic, machine to machine (M2M) communication, and Internet Technologies. Many researchers investigate to manage energy that controls thermal appliances in smart buildings, such as heating and air conditioning systems. One of the studies [82] distributed a simple and efficient energy management system by applying a novel approach. Develeped design have three main parts. The first is that collecting data from sensors “Temperature (T) and Humidity (H)”, then, monitoring the entire system over the cloud, and the

last, inspect devices “Air conditioners (AC)” by manipulating the level of fan speed, temperature, and the mode/state “Heat, Cool or Dry”. The aim of this study is to manage energy without affecting the comfort level of occupants. They prove that the system developed can reduce energy consumption automatically and maintain steady comfort level required by the occupants in the building.

Monitoring and maintaining air pollution level are of great importance in smart city management. The exordium of gases or aerosol particles emanate in the air stimulates direct or indirect damage to humans, plants, animals and other habitats of ecosystem and they give rise to serious health and environment threats. Global warming is also the kindling problem of the environment. Because of this condition, environmental parameter monitoring are needed in modern megalopolis due to revolution and advancement. The developed system perform Internet of Things (IOT) that provides an economical and an effective system to monitor air pollution level in particular area. IOT based monitoring devices defines the levels of toxicity in gases in the atmosphere. Air pollution can be defined by performing different sensors and GPS connected to ESP8266. The developed system help control the pollution level [83].

Since IoT data comes from heterogeneous sources having different format which is difficult to understand, therefore semantic web technologies have been applied to IoT for interoperability in order to reuse data among heterogenous IoT applications. Barnaghi et al. [44] present the pioneering work on semantics for IoT by reviewing recent developments on applying semantic technologies to IoT, e.g., information modelling, ontology design and processing of semantic data. Other recent works on interoperability in IoT are discussed in [45], [46], [47], [48], [49], [50], [51] and [52]. Although very good starting points have been identified, the area is still open to many challenges.

2.2.4 “SOTA” IN SMART CITY PLATFORMS

Smart City systems proposes for a wide range of city services such as transportation, traffic, air pollution, waste management, safety, energy, etc. Integrating these domains into a complete and consistent solution requires basic services provided by the underlying software infrastructure. A novel, comprehensive software platform could provide such services, including facilities for application development, integration, deployment, and management to ease the construction of sophisticated Smart City applications.

In recent years, the emergence of new information and communication technology has contributed to major changes in society. The growth of smart ICT devices changes the communication habits of people, their connection with the environment and with others, so that it creates an interconnected society. The development and structure of cities are affected by ICT and can be used to make them more sustainable. Social networks are a good example of this evolution. In general, a Social Networks Site (SNS) is defined as a service that enables users to share their personal profile and opinions regarding different topics. Social networks are also considered to be a powerful platform for acquiring knowledge of a city. They are currently one of the key points of intelligence provided to urban planners in order to figure out how public infrastructure is being used by people. Understanding the actual uses of urban public spaces plays an important role in the planning and

design of smart cities. Therefore, governance must take advantage of new opportunities provided by social networks to allow people to take part in the decision-making process. This causes administrations to step into citizen-centric management where government and citizens work together to build and design better cities. This issue has received less attention in the state-of the-art smart city platforms and PLODER aims to highlight this in the smart city platforms.

Smart City Platforms

There are various open source and commercial IoT platforms which have different architectures. The most cited smart city platforms are listed below:

- FIWARE: The European-funded IoT open source platform FIWARE has matured significantly in the past two years according to developers, and is now being used in industrial production cases, pilot smart cities, and utilities projects. Two projects using the FIWARE platform include a city water quality pilot and an early warning system to identify and prevent pest risks to agricultural crops. Advice company currently runs a smart city project in Seville, Spain, demonstrating how to monitor water quality at the city's public fountains.[61].
- EPIC: The European Platform for Intelligent Cities (EPIC) project proposes a complete IoT Middleware to facilitate the use and management of Wireless Sensor Networks (WSN). This middleware aims to deal with the heterogeneity, interoperability, scalability, extensibility, and reconfigurability problems in a WSN [62].
- “The ClouT project” (2013 – 2016) has provided a smart city platform which can gather data from heterogeneous data sources in a city (IoT devices, social networks, mobile applications, web data, etc.) and provide access to the data (and to actuators) in a unified way. The project has organized deployments in 4 pilot cities of the project: Santander, Genova, Fujisawa and Mitaka [63].
- Intelligent City Platform, which talks to the Cambridge’s existing “LoRaWAN” Internet of Things network centralizes the data gathered from sensors already deployed around the English city, ready for later analysis. The idea is that real-time data is available through the platform for developers and the like to build apps and solve problems. Examples suggested by the Smart Cambridge organization include predicting bus arrival times - a handy use case, as Londoners used to the metropolis’ “iBus” system of roadside beacons and sensors on buses will attest to [64].
- FESTIVAL is an H2020 EU-Japan collaborative project (2014 – 2017) that aims to federate heterogeneous testbeds, making them interoperable and building an “Experimentation as a Service” (EaaS) model. Going beyond the traditional nature of experimental facilities, FESTIVAL testbeds have heterogeneous nature. The FESTIVAL platform will be tested on three different smart city domains: smart energy, smart building and smart shopping [65].
- RERUM is an EU FP7 project, which is developing, evaluating, and demonstrating an architectural framework for reliable, resilient and secure IoT for Smart City applications. The framework is based on the concept of security and privacy by design, addressing the most critical factors for the success of Smart City applications. [66].

- OpenIoT: is an open source platform, handling almost all the main requirements like use of an IoT middleware to configure and collect data from devices, the middleware to store the data collected from sensors and the development tools. But, its architecture does not consider social networks, and does not provide support for pre-processing services relevant when dealing with Big Data [62].
- CiDAP The City Data and Analytics Platform is a Big Data–based platform that aims to use the city data collected from the city to enable context-awareness and intelligence in applications and services. This platform processes large datasets collected from an IoT Middleware. The strong points of its architecture are data storage and processing, real-time and batch processing modules. An important limitation of CiDAP is that the platform does not foresee specific services and tools for application developers and does not allow the deployment of new services in the platform, making its extensibility difficult [67].

Other urban data platforms and vendors also existing in the market are listed below:

- iSMOV, powered by Indra Sistemas S.A., one of the largest companies in its field inside the Spanish market. This system is focused on the management of transport, both public and private, within a specific population. Within iSMOV, various interrelated subsystems can be found, the objective of which is to optimize the public transport service, such as buses, taxis, bicycles, etc., thus reducing the need for citizens to use other private transport methods, along with reducing emissions of greenhouse gases, and noise.
- FIWOO, promoted by several Spanish corporations, such as Emergya, which offers digitization services internationally, and Secmotic, specialized in offering digital transformation solutions for cities, industry and businesses.Based on open source technologies, FIWOO is focused on managing the information generated by multiple Internet of Things (IoT) devices, thus improving the understanding of various systems distributed by a population. This platform is based on FIWARE technology, which is the standard promoted by the European Union related to the concept of smart city.
- Cisco Kinetic for Cities offers one of the most comprehensive smart city tool suites available on the market. Offered by the Cisco company, it adds an intelligent lighting system, which saves energy and money, while simplifying its maintenance. It also offers another module related to the mobility of transport through the city, which allows the management of parking lots and public roads. The monitoring and security system can detect any type of incident or theft that is perpetrated in the city. Finally, one of the most interesting modules that it offers is that of intelligent waste management: it allows to know in real time the state of the bins distributed throughout the city.
- The platform offered by Siemens is recognized for being one of the most complete and functional of the market. One of its main points is the air quality control module. This system is capable of predicting the quality of the breathable air in the coming days with a high hit rate, analyzing the data generated by various IoT sensors distributed throughout the city. In this way, Siemens is committed to improving the environment, through the digitization of information from cities.

The main hub of this platform is also very powerful. It shows in a very visual way the information collected by the different IoT devices in the city. It should be mentioned that Siemens is not in charge of the hardware devices themselves, only the collection and display of the data generated by them.
One of the use cases for this platform is inside Hong Kong. There, a digitized parking system has been implemented, making use of blockchain technology, to improve the security and reliability of this system. Cameras and other sensors are used at the same time.

- ESRI has become a major player in the platform for smart cities solutions because of their extensive presence in municipalities as a geospatial information system. On top of that, ESRI grew their proprietary platform allowing their partners to build their own solutions. The solutions based on Esri ArcGIS technology supports advanced automation and real-time integration of city monitoring and management systems. The solution offering varies from Geo database creation in 2D and 3D format to integrate these solutions with Smart Applications and devices for Spatial Decision Support and Action.
- The Japanese company NEC offers a smart cities platform called 'CCOC' (Cloud City Operations Center). This system offers a multitude of modules, such as the intelligent management of various basic public services, such as water and electricity, as well as public mobility, among others.
- One of the most outstanding subsystems in this platform is the automation of industries (Industries 4.0), which allow companies to be advised, based on data collected from their production plants, in order to optimize them to the maximum. Another of the most important modules is intelligent smart 'social security'. With this, it is intended to achieve a better and more personalized attention to the individual in health and wellness issues, from the information collected by other subsystems on each of the citizens.

  The most striking case of CCOC platform implementation is that of Lisbon, the capital of Portugal. Many of the subsystems offered have been implemented here, in relation to the management and administration of essential public services. As a feature, an intelligent video system has been developed, by which problems or incidents on public roads can be identified, making use of artificial intelligence technologies.

- Schneider Electric has built the smart cities platform focused on environmental sustainability, called 'EcoStruxure'. It has several modules, oriented to different places, such as the home, industry and other infrastructures. Without a doubt, the most striking subsystems are 'Building' and 'Grid.

  EcoStruxure Building is oriented to high-volume buildings and constructions. It allows saving on various resources, making smart measurements, and applying said information to electricity and water consumption, among others. EcoStruxure Grid enables power companies to deploy smart systems that integrate renewable energy sources into the grid easily, automatically and smartly.

- The Panasonic platform is called 'CityNOW', and it is geared towards the US market. Several important use cases of this platform can be found throughout the world. One of them occurs in Denver, Colorado. The aim is to implement an automated and safe road and circulation system,

where traffic accidents do not occur. This will also be achieved by connecting all the vehicles in circulation, so that they can communicate with each other, in order to avoid problems on the roads.

Finally, another of Panasonic's innovation projects in this field is in Fujisawa, Japan. Unlike in Denver, where this system is still under construction, in Fijusawa this platform has already been fully implemented. This project is focused on the modernization of a living area close to Tokyo, equipping it with intelligent resource management systems, such as water and electricity. The use of renewable energy is also encouraged, reducing CO2 emissions to the maximum.

- The Chinese company Huawei is a benchmark in mobile and communication technologies at a worldwide level. Within the smart cities market, it has various devices and platforms, where customers can build their own systems. It is important that Huawei does not offer specific services or modules like other platforms developed here, but rather offers its hardware technology and knowledge of mobile networks, 5G, IoT and cloud services to potential customers.

  Its platform has a main 'hub' of all the data generated by Huawei's own IoT devices, which are collected by the network deployed throughout the city of LTE and 5G connectivity, in which this Chinese company is a leader in terms of implantation. They also make a clear commitment to data governance with respect to the city's politics, through their ‘eGovernment’ platform.

- The smart cities platform developed by IBM uses its own cloud computing and artificial intelligence technologies as its foundation. It offers a multitude of services and solutions, such as the management of public and social services, such as the police or fire; account and tax administration to avoid fraud as much as possible.

  Two of the subsystems that stand out from this offer are national and international defence, focusing on border and customs control. With this system, more efficient control of training in the army can be carried out, as is the case in Sweden, where this system has been successfully implemented. The digitization of border and customs control allows the client to maintain a much more detailed and effective system management, while avoiding certain problems such as corruption and illegal traffic between countries.

  This platform has been implemented to a greater or lesser extent in various countries and cities. For example the case of Switzerland. It has implemented the border control module in its territory, to streamline the process of entry and exit of goods and people to a greater extent. Furthermore, the California state government has integrated these management and control systems of expenses and taxes in its administration to prioritize single payment houses, automatically determining which ones have the best chance of being resolved.

- Microsoft also offers intelligent systems related to the management of public transport, police security, etc. The health and social services module allow serving both citizens and insurance companies to improve the service provided to citizens. It can also be used for remote service provision.

Another of the most relevant subsystems of the Microsoft smart cities package is related to payment and tax management. Through the digital transformation of these services, and making use of the latest technologies, such as blockchain, it is intended to simplify the payment of taxes to the administration, while saving on expenses.

Many of the main cities in the world that are implementing this type of digital services are betting on Microsoft's proposal. They are the case of the city of Denver in EE. USA, and Taipei.

In Denver they have opted for the implementation of services to improve the lives of their citizens. Technologies such as Azure Data Lake and Azure IoT Hub, enable the ability to intelligently manage traffic signals to avoid traffic jams, and install devices that control air quality in urban areas.

On the other hand, in the city of Taipei, and with the help of AAEON as a partner, systems have been installed to improve the efficiency of electrical systems. Traffic lights and other traffic signs are an example of systems that have been digitized to improve traffic management. The consumption of energy by public lighting has also been reduced, making it a more intelligent use of it. The next phase of this plan is the implementation of a 5G network throughout the city, to manage all these IoT devices more efficiently.

- CivicConnect offers its own SaaS (Software as a Service) over Amazon's infrastructure, which includes applications for different platforms, for both users and administrators. They offer a virtual reality service for compatible devices, through which you can obtain more information about the city, making use of the latest technologies.

  There is also an offer from other companies, such as Mutualink, which offers intelligent services related to health and social security; Accelerite has a smart cities platform designed for people without computer knowledge, where ease of use and simplicity is the main thing.

- The SAP company has built quite a few 'smart city modules', such as smart financial control, smart roads & improved public services, like health and security. All of this has been built on top of its data-base and cloud infrastructure, called 'SAP S/4HANA'. This tool serves as a dashboard for the city council, to manage and easily visualize all concurrent data flows in the city, in a single screen.

  This firm has a lot of past experience in smart city projects, having developed strong tools and a unified team. The most successful use case of this is present inside the Auckland City, USA.

  In this case, the SAP platform helped Auckland to amalgamate eight local governments into only one, improving most services provided by the administration in the process. They now can generate useful information more reliably, based on all the raw data they generate in that huge city.

Comparative table

After presenting all details related to the most important smart city platforms available in the market at this moment, a more structured comparative analysis of all of them will be performed.

Platform	FIWARE powered	Smart grid	Smart transportation	Smart industry	Smart services
iSMOV	No	Yes	Yes	No	Yes
FIWOO	Yes	No	Yes	Yes	No
Cisco KINETIC	No	Yes	Yes	Yes	Yes
Huawei	*Only serves as a hardware platform				
IBM	No	Yes	Yes	Yes	Yes
Microsoft	No	Yes	Yes	Yes	Yes
Amazon	*Only serves as a software platform				
SAP	No	Yes	Yes	Yes	Yes
Esri	No	Yes	Yes	No	No
Siemens	No	Yes	Yes	Yes	No
Hitachi	No	Yes	Yes	Yes	Yes
NEC	Yes	Yes	Yes	Yes	Yes
Schneider ecoStruxure	No	Yes	No	Yes	No
Panasonic CityNOW	No	Yes	Yes	No	No

Table 2 Smart City Platforms in the market

As you can see in the table above, just two of all studied platforms are compatible with the FIWARE platform, a feature that could become pretty much essential in the future, specifically in the European market.

Another piece of information that can be extracted from the table is that both smart grid and transportation systems are key in most current smart city platforms, as almost all of them offer this kind of systems. Other modules such as smart industry and services appear to be less important for both users and the industry, but many of the important ones like IBM, Hitachi or NEC offer these systems, among other one, as a key differentiator against the competition.

Challenges

Santana et al. [62] have analysed 23 smart city platforms and enumerated the most frequently cited open research challenges and problems in the development of next generation, robust software platforms for smart cities:

- **Privacy**: Protecting data collected from citizens, city, and enterprises,
- **Data Management**: Managing all the data collected in the platform,
- **Heterogeneity:** Ensuring the interoperability of devices and applications for different domains,
- **Energy Management**: Managing the electricity used by devices deployed in the city,
- **Communication**: Enabling communication among heterogeneous devices,
- **Scalability**: Allowing the growth of devices and users connected to the platform,
- **Security**: Protecting the city data, services, and infrastructure,
- **Lack of Testbeds**: There are insufficient testbeds to experiment Smart City solutions,
- **City Models**: Defining a model describing the city. Semantic Web and Ontologies,
- **Platform Maintenance**: Maintaining the city systems and infrastructure.

The most cited problem in the literature is that of ensuring the privacy of user data, because of the amount of personal and critical data that a platform needs to handle (special concerns are user locations and medical records)

The second most cited challenge is heterogeneity, because of the large number of different systems, services, applications, domains and devices that a platform must support.

An important and understudied issue is how to create a generic platform to support different cities' requirements. We should consider that each city has unique characteristics and size and smart city platforms need to be adapted to fit these diversities.

As part of the EU smart city project RUGGEDISED and in collaboration with the EU SCC-network of smart city projects and the EIP Smart Cities marketplace, the Erasmus University Rotterdam researched and gathered data from more than 100 respondents in 80 European cities, most municipality staff responsible for urban data platform development, and 85% were partners in one of the EU SCC Lighthouse projects, funded by the European Commission. The study concluded in mid-January 2020.

The study analysed the stage of development on urban data platforms; the vision behind these platforms; the business and technology design; the implementation barriers and accelerators, and the use and impact of these platforms. The study aims to share learnings on use cases for data management of urban data platforms among European smart cities.

Key findings from the study included five relevant aspects for the POLDER project summarized below:

- Regarding the purpose and scope of Urban Data Platforms, the top motives and ambitions for UPDs are **better policy and decision making**; and **cost efficiency and operational**, while **data privacy and security** underpin every other as top priority.
- 50% of the cities surveyed have clear ambitions to establish an **open and interoperable data platform** that encompasses the entire city and supports multiple systems. In this way, they

want to try to achieve a certain independence grade of the platforms and systems in use, in case they want to change services in the future. Architecture choice is primarily driven by the requirements to facilitate openness of the platform, without **risk of vendor lock-in**. The use of open standards supports this.

- Regarding the financial risks linked to the implementation of these kinds of platforms in the city, the most important factor is to develop a **real & sustainable set of use cases**. Those will drive innovation forward, developing trust with the implementer of the platform and investors in the way. But also, the **time and price of implementation** of these services must be substantially improved, making them more attractive to potential investors.

2.2.5 "SOTA" IN NETWORKING & DATA SECURITY

The convergence of both Network Connectivity as a Service and Network Security as a service has been named by GARTNER as SASE "Secure Access Service Edge". POLDER project does not aim to develop a SASE solution but will take advantage of Data In Motion Security thanks to SMP (Simultaneous Multi Path) and will consider SASE as a Smart City communications enabler.

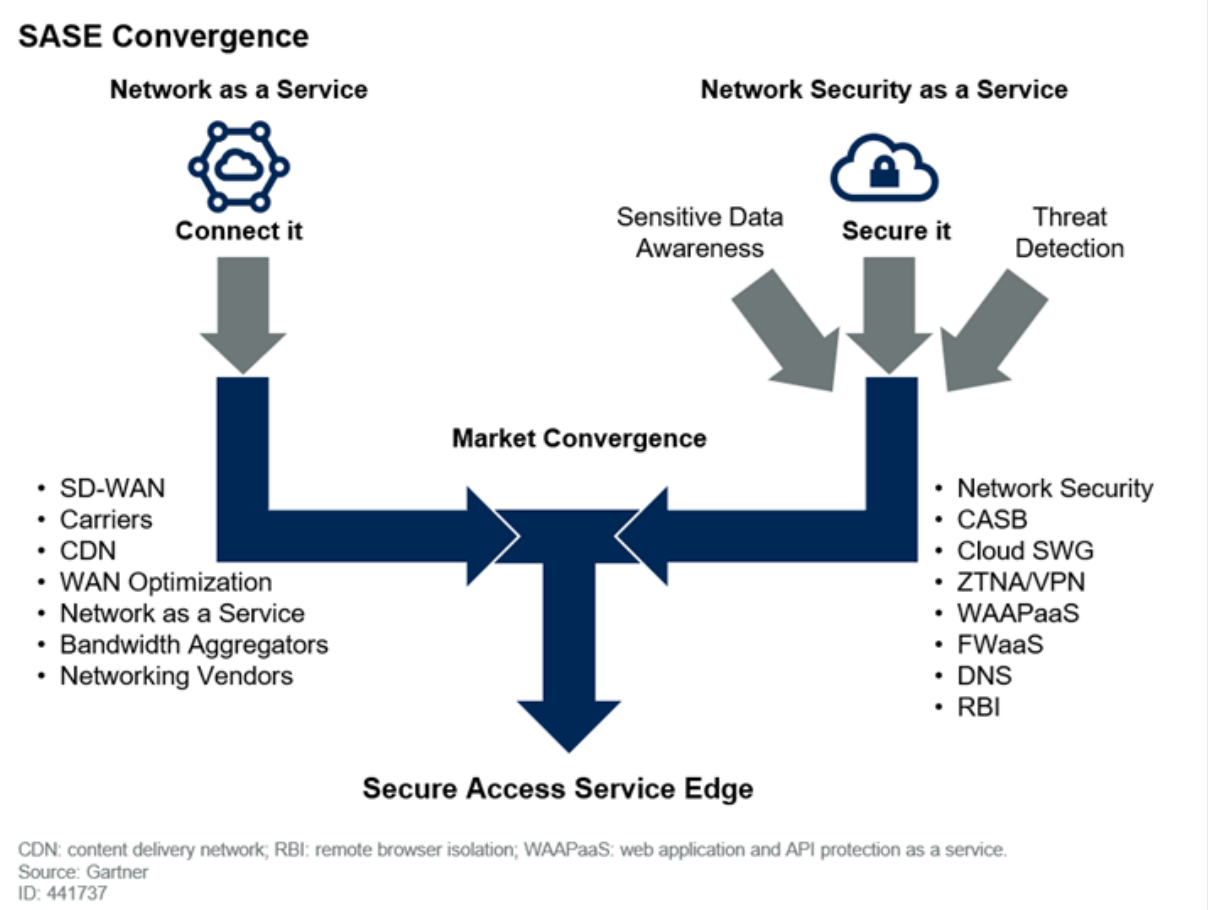

Figure 1 Secure Access Service Edge

2.2.6 "SOTA" IN DATA GOVERNANCE

Data Governance is the process, and procedure organizations use to manage, utilize, and protect their data. In this context, data can mean either all or a subset of an organization's digital and/or hard copy assets. In fact, defining what data means to an organization is one of the data governance best practices.

We understand data governance as the who, what, when, where, and why about an organization's data.

Another aspect of data governance is protecting the Smart City and citizens private data. Data breaches are near-daily occurrences in 2019 and governments are enacting laws – HIPAA, GDPR, CCPA, and more – to protect the private data of citizens. A data governance program builds controls to protect data and help organizations adhere to compliance regulations.

2.2.7 "SOTA" IN DATA VISUALISATION

There are so many popular data visualisation tools in the market, like Apache Zeppelin, Elastic.co Kibana or Grafana. Furthermore, there are very mature business intelligence tools like Tableau, Microsoft Power BI or QlikView. Since POLDER backend services will be standards compatible, any of these tools can be adopted for visualisation. Since there are so many alternative tools, for practical reasons this section will focus the solutions in the FIWARE software stack which are Wirecloud and Knowage.

Wirecloud

Wirecloud is a web application mashup platform that aims to allowing end users without programming skills to easily create web applications and dashboards. It has default data source connectors for CKAN and Orion Context Broker. In Wirecloud, workspaces are the pages for group of operators and widgets. Widgets provide user interface for any purpose (eg. Showing data on map, showing a list, checkbox etc.), on the other hand, operators provide backend jobs (eg. Getting data from a data source, filtering, selecting specific values etc.). Operators and Widgets can be connected in the wiring page as shown below. (Green = Operator, Orange = Widget)

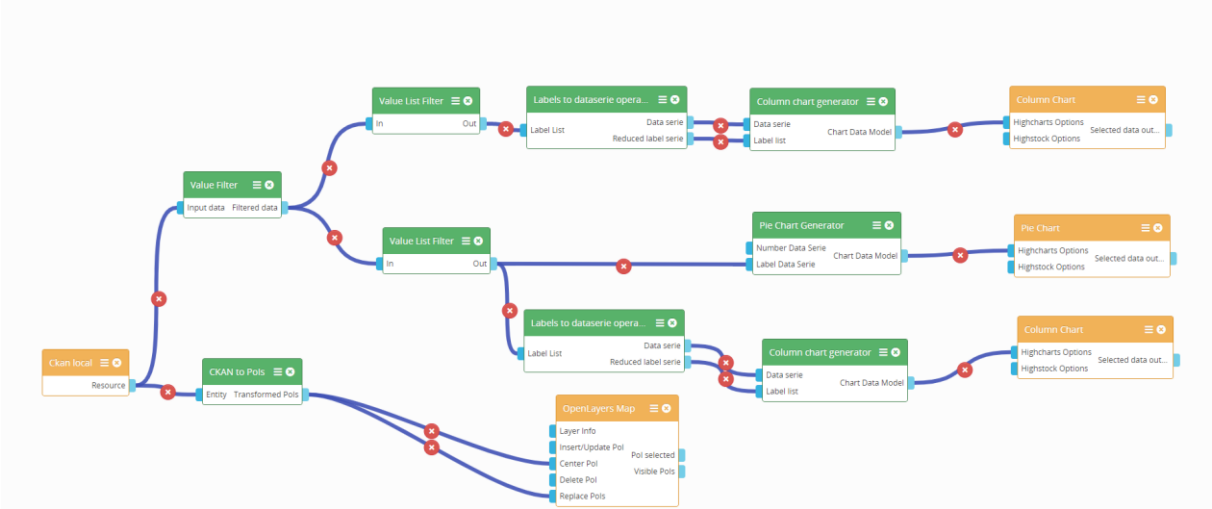

Figure 2 Wirecloud Operators and Widget Wiring

Knowage

Knowage is an open source Business Analytics suite that composed of several modules, each module conceived for a specific analytical domain. Knowage provides advanced self-service capabilities that give autonomy to the end-user, able to build his own analysis, explore and organize data. Knowage can be connected to various data sources (HDFS, PostgreSQL, MySQL, etc.), and provides a query interface to pull data from any source with given filters, and in given shape. That pulled data can be used in various visualize tools like bar charts, pie charts, map, table, etc. Cockpit is the general name for these visuals. An example of cockpit is given below.

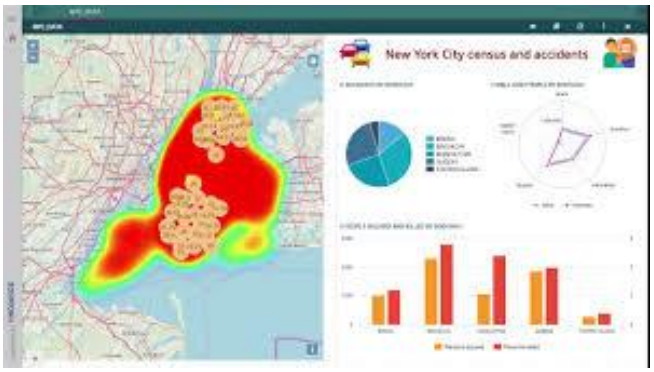

Figure 3 An Example view from the Knowage Dashboard

2.3 SotA in collaborative research

Link to previous and/or current collaborative research projects:

Project Name	Coop. Programme	Time period (approx.)	Technical Focus	Relationship
RUGGEDISED	H2020	2015-2019	Energy, e-mobility, thermal, electricity	City planning modules (urban strategy system)
CAP	ITEA	2013-2016	Collaborative analytics platform	Analytics platform, using data from multiple data sources, data business challenges
FUSE-IT	ITEA	2014-2017	Smart building management	Managing energy aspects of a building and aggregation of sensor data from various domains
TESTOMAT project	ITEA	2017-2021	Test automation	Data platform development
Personal Health Empowerment	ITEA	2017-2020	Integrator, Coaching Tool Developer, Software Engineering	Empowering people to monitor and improve their health using personal data and digital coaching. Sensor based data is collecting from wearing devices.
Logistical Warehouses Automation, Communication and Information Infrastructure	Republic of Turkey Prime Ministry Disaster & Emergency Management Authority (AFAD)	2014-2017	Gathering data, apply Machine Learning Machine Learning and Deep Learning techniques for suitable warehouse identification and monitoring according to temperature, humidity and coordinates	Management and tracking of the inventory in 28 logistical warehouses throughout Turkey
REMOURBAN	H2020	2015 - 2020	A sustainable urban regeneration model leveraging the convergence	Eskisehir, Turkey is one of the lighthouse cities.

PROJECT NAME	COOP. PROGRAMME	TIME PERIOD (APPROX.)	TECHNICAL FOCUS	RELATIONSHIP
			of energy, mobility and ICT to transform European cities into Smart Cities	
OASIS	H2020	2016-2018	Linked Data for more discoverable public services, public services, smart cities	citizens facilitation
C3PO	ITEA	2014-2017	Focus on city planning (urban areas, buildings), to bring different stakeholders and information together.	Managing with data from multiple sources
PS-CRIMSON	ITEA	2016-2019	Combining information from different silos in an integrated model	Especially the platform architecture and data fusion is of very high interest to POLDER (which goes further in policy support)
PARFAIT	ITEA	2017-2020	Personal data protection for IoT data	Privacy of data from IoT sensors
CityStory	ITEA	2018-2020	Creating ways to involve citizens with smart city issues, through innovative media concepts	See note

Table 3 Related collaborative research projects

Note: The ITEA "CityStory" project is expected to run in parallel with POLDER. The goals of "CityStory" link well to the POLDER goals, while concentrating on different parts of the technology chain. "CityStory" focuses on the communication between Smart City and citizen. The insights and results of "CityStory" can be taken up by POLDER in order to improve the way how POLDER communicates with citizens; alternatively, POLDER may provide use cases for "CityStory". Especially WP7 (Decision making) could benefit from collaboration with "CityStory". The projects share some countries and even one pilot location (Antwerp in Belgium). Contacts with the "CityStory" coordinator have been established and will be followed up once both projects are starting.

“SotA” in Data management versus Data strategy framework”.

Data Management is: “The development, execution, and supervision of plans, policies, programs, and practices that deliver, control, protect, and enhance the value of data and information assets throughout their lifecycles.”

Data Strategy: “Typically, a Data Strategy requires a supporting Data Management program strategy – a plan for maintaining and improving the quality of data, data integrity, access, and security while mitigating known and implied risks. The strategy must also address known challenges related to Data Management.”

3. Standards

Standards help create the foundation for interoperability where next-generation technologies and capabilities can be implemented and aligned cost-effectively and smoothly. Main standardization organizations have published standards related to smart city platforms in different points of view because of breadth and scope of Smart City activities.

There are many working groups and bodies working on standard, model, infrastructure and evaluation methods on the concept of smart cities. The most cited of these are listed below.

- ITU-T Study Group 5 – Focus Group on Smart Sustainable Cities, aims to act as an open platform for smart-city stakeholders “Smart Sustainable Cities’’ is also the theme of ITU’s 3rd Green ICT Application Challenge [60].
- IEC launched Systems Evaluation Group (SEG) on Smart Cities to identify standardization areas that address Cities’ relevant cross-cutting issues [70].
- IEEE has been working for many years on the infrastructure, networking, generation, automation, operation, and distribution necessary to design, support, deliver, and connect new energy to the cities and homes that demand it. There are over 150 resulting active standards and standards’ projects for Smart Cities [71].
- European Commission: By launching the Smart Cities and Communities European Innovation Partnership (SCC), the European Commission aims to boost the development of smart technologies in cities.
- NIST established the cyber-physical systems public working group (CPS PWG) to work on security phase of smart cities [73].
- ANSI convened a Joint Member Forum with subject matter experts from standards developing organizations, industry, government, and academia to discuss the role that standards and conformance solutions can play in contributing to national and international Smart City initiatives.
- BSI develops strategies which aims to accelerate the implementation of Smart Cities and minimize the risks of failure. The strategy outlines a foundation of knowledge to help cities as the embark on a program to become smarter [72].

3.1 Data

Below table[1] lists the standards related to the data aspects of a smart city. The table is filtered and simplified in order to match smart city domain.

Activity domain	Entity	Title of deliverable	Scope of deliverable	Current status
Geospatial Information	ISO	ISO 19112:2003 Geographic information – Spatial referencing by geographic identifiers	ISO 19912:2003 defines the conceptual schema for spatial references based on geographic identifiers. It establishes a general model for spatial referencing using geographic identifiers, defines the components of a spatial reference system and defines the essential components of a gazetteer. Spatial referencing by coordinates is not addressed in this document; however, a mechanism for recording complementary coordinate references is included.	International Standard
Geospatial Information	OGC	KML	KML is an XML language focused on geographic visualization, including annotation of maps and images. Geographic visualization includes not only the presentation of graphical data on the globe, but also the control of the user's navigation in the sense of where to go and where to look.	Implementation Standard
Geospatial Information	OGC	Web Map Service (WMS)	The Web Map Service Interface Standard provides a simple HTTP interface for requesting geo-registered map images from one or more distributed geospatial databases. A WMS request defines the geographic layer(s) and area of interest to be processed. The response to the request is one or more geo-registered map images (returned as JPEG, PNG, etc) that can be displayed in a browser application.	Implementation Standard
Geospatial Information	OGC	Geography Markup Language (GML)	GML is an XML grammar for expressing geographical features. GML serves as a Modelling language for geographic systems as well as an open interchange format for	Implementation Standard

[1] https://www.itu.int/en/ITU-T/jca/iot/Documents/deliverables/Free-download-IoT-roadmap.doc

Activity domain	Entity	Title of deliverable	Scope of deliverable	Current status
		Encoding Standard	geographic transactions on the Internet. As with most XML based grammars, there are two parts to the grammar – the schema that describes the document and the instance document that contains the actual data. A GML document is described using a GML Schema. This allows users and developers to describe generic geographic data sets that contain points, lines and polygons. However, the developers of GML envision communities working to define community-specific application schemas that are specialized extensions of GML. Using application schemas, users can refer to roads, highways, and bridges instead of points, lines and polygons.	
Geospatial Information	OGC	City Geography Markup Language (CityGML) Encoding Standard	This document is an OpenGIS® Encoding Standard for the representation, storage and exchange of virtual 3D city and landscape models. CityGML is implemented as an application schema of the Geography Markup Language version 3.1.1 (GML3). CityGML models both complex and georeferenced 3D vector data along with the semantics associated with the data. In contrast to other 3D vector formats, CityGML is based on a rich, general purpose information model in addition to geometry and appearance information. For specific domain areas, CityGML also provides an extension mechanism to enrich the data with identifiable features under preservation of semantic interoperability.	Implementation Standard
Geospatial Information	OGC	Web Feature Service	This International Standard specifies the behaviour of a service that provides transactions on and access to geographic features in a manner independent of the underlying data store. It specifies discovery operations, query operations, locking operations, transaction operations and operations to manage stored parameterized query expressions.	Implementation Standard

Activity domain	Entity	Title of deliverable	Scope of deliverable	Current status
Geospatial Information	OGC	OGC Filter Encoding Encoding Standard	This International Standard describes an XML and KVP encoding of a system neutral syntax for expressing projections, selection and sorting clauses collectively called a query expression. These components are modular and intended to be used together or individually by other standards which reference this International Standard.	Implementation Standard
Context Information Management	ETSI ISG CIM	ETSI GS CIM 009 V1.1.1 (2019-01) 'Context Information Management (CIM); NGSI-LD API'	This Group Specification provides additions and corrections to the GS-004 prelimAPI specification, based on feedback about GS-004 received from other SDOs as well as developers in the linked-data, internet-of-things, and mobile-apps and smart-applications communities, as well as from end users and stakeholders. In particular open issues and proposed features in Annexes of the GS-004 document, including the topics of privacy/security, will be considered for inclusion as part of this specification.	Published
Context Information Management	ETSI ISG CIM	ETSI GR CIM 002 V1.1.1 (2018-09) 'Context Information Management (CIM); Use Cases (UC)'	The purpose of this Group Report is to collect and analyse use cases and requirements for management of information coming from many different sources (not only IoT) and data models to allow applications to perform updates on context, register context providers, query information on current and historic context information and subscribe for receiving notifications on context changes. Wherever possible, material shall be collected (by reference) from prior art and existing standards and specifications in this area.	Published

Table 4 Data Standards in Smart Cities

3.2 Internet of Things

IEEE 2413-2019 (IEEE Standard for an Architectural Framework for the Internet of Things (IoT)) provides an architecture framework description for the Internet of Things (IoT) which conforms to the international standard ISO/IEC/IEEE 42010:2011 is defined. The architecture framework description is motivated by concerns commonly shared by IoT system stakeholders across multiple

domains (transportation, healthcare, Smart City, Smart Grid, etc.). A conceptual basis for the notion of things in the IoT is provided and the shared concerns as a collection of architecture viewpoints is elaborated to form the body of the framework description. Standards Committee: BOG/CAG - Corporate Advisory Group, Published Date: 2020-03-10; https://standards.ieee.org/standard/2413-2019.html)

- **P2413.1 - Standard for a Reference Architecture for Smart City (RASC)**

This standard provides an architectural blueprint for Smart City implementation leveraging cross-domain interaction and sematic interoperability among various domains and components of a Smart City. This standard also leverages an architectural framework for the IoT defined in the draft of IEEE P2413 standard, which relies on the international standard ISO/IEC/IEEE 42010. Based on requirements from government bodies, enterprises, and consumers, a variety of applications will be delivered to create social value.

The Smart City applications include water management, waste management, smart streetlights, smart parking, environment monitoring, smart community, smart campus, smart buildings, eHealth, eLearning, eGovernment, etc.

This standard defines;

- a Reference Architecture for Smart City that includes Smart City Intelligent Operations Center (IoC), Internet of Things (IoT), including descriptions of various IoT vertical applications in Smart City, and identification of commonalities between different vertical applications in Smart City.
- the four layers of the Smart City architecture, device layer, communication network layer, IoT platform layer, and application layer.
- relationships with and attributes specific to the cloud computing center, the edge computing technologies, and big data analysis related to IoT for Smart City.

The standard describes;

- the Intelligent Operations Center (IOC) which aggregates a wide range of data to visualize the city operational status, enables efficient collaboration across agencies and applications, and facilitates decision-making based on knowledge derived from Big Data.
- unified security aspects of the Smart City architecture.

3.3 Ontologies

Ontologies can be classified according to the scope of the objects described by the ontology. For instance, the scope of a local ontology is narrower than the scope of a domain ontology; domain ontologies have more specific concepts than core reference ontologies, which contains the fundamental concept of a domain. Foundational ontologies can be viewed as meta ontologies that describe the top-level concepts or primitives used to define other ontologies. Finally, general ontologies are not dedicated to a specific domain thus its concepts can be as general as those of core reference ontologies.

Core ontology defines concepts which are generic across a set of domains. This type of ontology is linked to domain ontologies, but it integrates different viewpoints related to specific group of users. This type of ontology is the result of the integration of several domain ontologies. A core reference ontology is often built to catch the central concepts and relations of the domain.

Domain ontology is only applicable to a domain with a specific viewpoint. That is to say that this viewpoint defines how a group of users conceptualize and visualize some specific phenomenon. The main domains that will be covered in POLDER are:

- Geographical Information Systems (GIS)
- Building Information Modelling (BIM)
- Traffic and transportation
- Air quality
- Wastewater
- Energy
- Waste Management
- Urban Planning

3.3.1 GIS

The City Geography Markup Language (CityGML) is an open data model for the storage and exchange of virtual 3D city models [77]. The standard is subdivided into several modules. Beside the Core module, 13 thematic extension modules are available in CityGML 2.0: Appearance, Bridge, Building, CityFurniture, CityObjectGroup, Generics, LandUse, Relief, Transportation, Tunnel, Vegetation, WaterBody and TexturedSurface.

The CityGML Building model is an excerpt of the CityGML standard in which only the concepts related to IFC are represented.

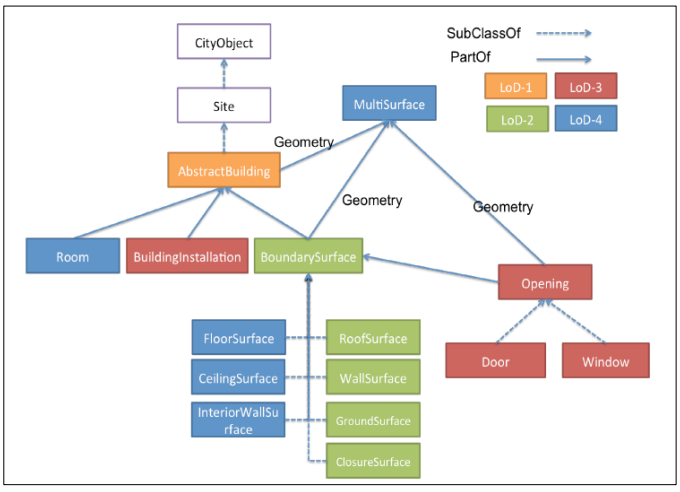

Figure 4 CityGML and Building Information

3.3.2 TRANSPORT

The core ontology covers different existing domain ontologies in the domain of transport and traffic:

- Ontology of Transportation Networks (OTN), defined within the framework of the REWERSE project [78], describing various transportation concepts but not related to soft mobility.
- Ontology of Urban Planning Process (OUPP), including concepts representing soft mobility aspects [76]
- Datex II: standardised way of communicating and exchanging traffic information between traffic centres, service providers, traffic operators and media partners [75].
- CityGML for using the geometrical representations of objects and linking OUPP, OTN and Datex with other domain ontologies.

OTN

The ontology of Transportation Networks (OTN) has been defined within the framework of an European project REWERSE. OTN describes different concepts related to transportation, and with a focus on urban transport. It doesn't include concepts related to soft mobility. The main concept of OTN is Route of a Public_Trapsort that contains Route_Section. Route_section are defined based on 2 geometrical representations: Start_Point and Stop_Point. For representing different types of urban transport, OUPP allows the generalization of concepts like Route_Sections and Routes as general concept and to associate them with an attribute indicating the type of urban transport (soft mobility, public transport or other means of transport).

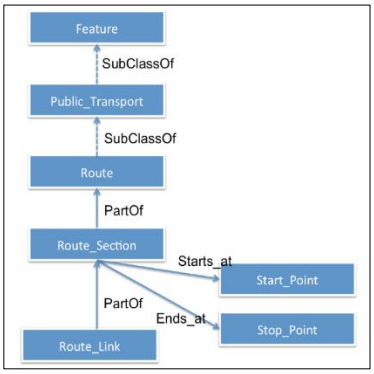

Figure 9. *Ontology of OTN (main concepts)*

OUPP

OUPP provides an integrated view of different aspects related to soft mobility. The major aspects defined in OUPP are transport route, cycle route and pedestrian route.

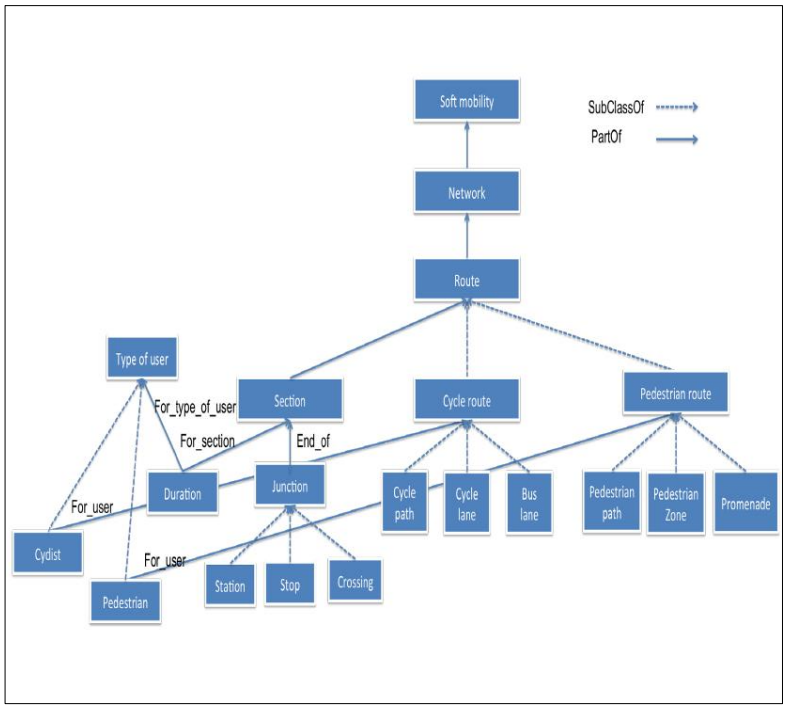

Figure 5 OUPP: part of the ontology of soft mobility

CityGML

CityGML has a transportation module that covers all kinds of traffic infrastructure except bridges, tunnels and waterways, which are modelled in separate modules. The main feature TransportationComplex represents for example roads or railways. The representation of a TransportationComplex starts with the Level of Detail 0 modelling of a linear network. With LoD 1 the TransportationComplex provides surface geometry. In the higher Levels of Detail (LoD 2, LoD 3 and LoD 4) the TransportationComplex can be further subdivided into TrafficArea (e.g. driving lanes, bicycle line or pedestrian zones) and AuxiliaryTrafficArea (e.g. kerbstones or green areas) [77].

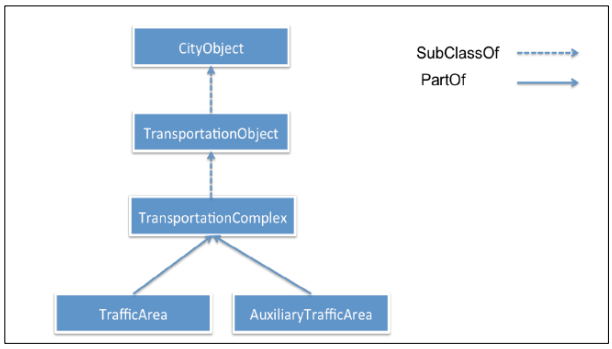

Figure 6 Transport in CityGML

3.3.3 BIM

IFC is a European research initiative for representing parts of buildings and their spatial relations. There is no universally accepted building model for IFC. Hence, the proposed IFC building model is defined based on IAI and ISO.

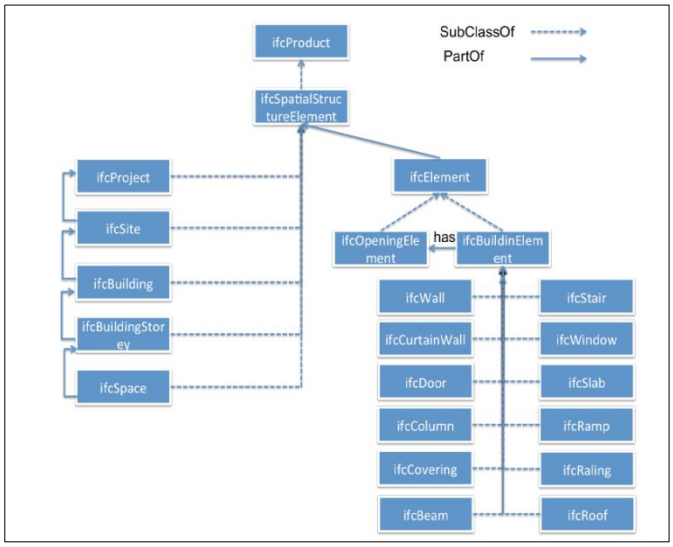

Figure 7 IFC: part of the ontology

3.4 KPIs/metrics for Smart Cities

Cities need indicators to measure their performance. A large variety of indicator frameworks exist to assess the sustainability and smartness of the cities. Related international standardization work has been carried out by three bodies, i.e. by **ISO** (International Standards Organization) and **ITU** (International Telecommunication Union) worldwide and by the coalition of the European standardization organizations **CEN** (Comité Européen de Normalisation- European Committee for Standardization), **CENELEC** (Comité Européen de Normalisation Electrotechnique- European Committee for Electrotechnical Standardization) and **ETSI** (European Telecommunications Standards Institute) in Europe. Currently, there are seven international city indicator standards

relevant for Smart sustainable city evaluation and reporting and this study focused on existing indicator frameworks for smart sustainable cities in the following published standards related to Smart City activities.

Standards developed by ISO Technical Committee 268 (ISO/TC 268) related KPIs:

ISO focuses on city outcomes to deal with challenges such as climate change, rapid population growth, and political and economic instability. The published standards regarding Smart City activities are as follows:

- **"ISO 37120:2018 Sustainable Development of Communities- Indicators for City Services and Quality of Life":**

  ISO 37120:2018 is focused on city services and quality of life as a contribution to the sustainability of the city. As part of a new series of International Standards being developed for a holistic and integrated approach to sustainable development, this standard includes indicators for city services and quality of life, indicators for smart cities and indicators for resilient cities.

  These indicators can be used to track and monitor progress on city performance. The indicators and associated test methods in this standard have been developed in order to help cities:

  a) measure performance management of city services and quality of life over time;

  b) learn from one another by allowing comparison across a wide range of performance measures; and,

  c) support policy development and priority setting.

  (https://www.iso.org/obp/ui/#iso:std:iso:37120:ed-2:v1:en)

- **"ISO 37122:2019 Sustainable cities and communities — Indicators for smart cities"**

  The indicators detailed in ISO 37120 have quickly become the international reference point for sustainable cities. ISO/TC 268/WG2 experts have identified the need for additional indicators for smart cities.

  ISO 37122 is intended to complement ISO 37120 and to establish indicators with definitions and methodologies to measure and consider aspects and practices that dramatically increase the pace at which cities improve their social, economic and environmental sustainability outcomes.

  When used in conjunction with ISO 37120, it is expected that this standard helps cities to identify indicators for applying city management systems such as ISO 37101 and to implement smart city policies, programmes and projects to:

  - respond to challenges such as climate change, rapid population growth, and political and economic instability by fundamentally improving how they engage society;

- apply collaborative leadership methods, work across disciplines and city systems;
- use data information and modern technologies to deliver better services and quality of life to those in the city (residents, businesses, visitors);
- provide a better life environment where smart policies, practices and technology are put to the service of citizens;
- achieve their sustainability and environmental goals in a more innovative way;
- identify the need for and benefits of smart infrastructure;
- facilitate innovation and growth;
- build a dynamic and innovative economy ready for the challenges of tomorrow.

(https://www.iso.org/obp/ui/#iso:std:iso:37122:ed-1:v1:en)

- **"ISO 37123:2019 Sustainable cities and communities — Indicators for resilient cities"**

  Just published (Publication date: 2019-12), "ISO 37123 *Sustainable cities and communities – Indicators for resilient cities"* is the first International Standard that provides a set of indicators on resilience by which cities can measure where they stand. ISO 37123 is also intended to complement other standards in the series on smart cities indicators that includes ISO 37120 and ISO 37122.

  A resilient city is able to prepare for, recover from and adapt to shocks and stresses. Cities are increasingly confronted by shocks, including extreme natural or human-made events which result in loss of life and injury, material, economic, and/or environmental losses and impacts. These shocks can include but are not limited to floods, earthquakes, hurricanes, wildfires, volcanic eruptions, pandemics, chemical spills and explosions, terrorism, power outages, financial crises, cyber-attacks and conflicts. A resilient city is also able to manage and mitigate ongoing human and natural stresses in a city relating to environmental degradation (e.g. poor air and water quality), social inequality (e.g. chronic poverty and housing shortages) and economic instability (e.g. rapid inflation and persistent unemployment) that cause persistent negative impacts in a city.

  The indicators in this standard have been developed to help cities:

  a) prepare for, recover from and adapt to shocks and stresses;
  b) learn from one another by allowing comparison across a wide range of performance measures, and by sharing good practices.

  The indicators in this document can be used to track and monitor progress towards a resilient city, through the development of a city resilience strategy or when applying a city management system such as ISO 37101.

  (https://www.iso.org/obp/ui/#iso:std:iso:37123:ed-1:v1:en)

Standards developed by ITU-T Study Group 5 – Focus Group on Smart Sustainable Cities:

ITU explicitly cites the idea of using ICT technologies to improve life quality, efficiency of urban operation and services, and competitiveness. ITU proposes a four-layer platform for smart cities:

- **Sensing Layer**, which consists of terminal nodes (sensors, actuators, cameras, and RFID readers) and capillary networks.
- **Network Layer**, which is the interconnection of the various city networks provided by the city's telecommunication operators.
- **Data and Support Layer**, which stores data collected by the city terminal nodes and data from other city services and applications. This layer is also responsible for providing services and data (raw and analyzed) to the city applications.
- **Application Layer**, which includes various services and applications available for citizens, managers, and companies [62].

A series of Recommendations of ITU give a general guidance to cities and provide the definitions of key performance indicators (KPIs) related to the use of information and communication technology (ICT) in the context of smart sustainable cities (SSCs). They specifically provide the KPIs related to ICT adoption and use in the context of SSCs. These are:

- **"ITU-T Y.4901/L.1601** Key Performance Indicators Related to the Use of Information and Communication Technology in Smart Sustainable Cities**"**
- **"ITU-TY.4902/L.1602** Key Performance Indicators Related to the Sustainability Impacts of Information and Communication Technology in Smart Sustainable Cities**"**
- **"ITU-T Y.4903/L.1603** Key Performance Indicators for Smart Sustainable Cities to Assess the Achievement of Sustainable Development Goals**"**

These Recommendations can be utilized by:

- Cities and municipal administrations, including the SSC-relevant policy-making organizations, and government sectors, enabling them to develop strategies and understand the progress related to the use of ICT for making cities smarter and more sustainable.
- City residents and their non-profit organizations, enabling them to understand the development and progress of SSCs.
- Development and operation organizations of SSC, including planning units, SSC-related producers and service providers, operation and maintenance organizations, helping them to fulfil the tasks of sharing information related to the use of ICT in the city.
- Evaluation agencies and academia, supporting them in selection of relevant KPIs for assessing the contribution from ICT in the development of SSC.

ETSI TS 103 463, Access, Terminals, Transmission and Multiplexing (ATTM); Key Performance Indicators for "Sustainable Digital Multiservice Cities":

ETSI TS 103 463 describes the selection of indicators for assessing indicators on city level. Starting from the definition of a smart city, indicators have been selected that can function as Key

Performance Indicators for tracking the progress towards city objectives. The indicator framework and indicators are based on CITYkeys deliverable 1.4 *(CITYkeys deliverable D1.4: CITYkeys indicators for smart city projects and smart cities- 24-01-2017).*

The indicators for smart cities focus on monitoring the evolution of a city towards an even smarter city. The time component -"development over the years"- is an important feature. The city indicators may be used to show to what extent overall policy goals have been reached, or are within reach.

With a starting point in the smart city definition, and taking into account the wishes of cities and citizens with regard to smart city indicators, the indicators are arranged in an extended triple bottom line sustainability framework, including the themes people, planet, prosperity, governance and propagation, and completed with specific smart city indicators.

Under the main themes subthemes conforming to major policy ambitions have been identified. Under these subthemes in total 73 city indicators have been selected. The selection has been based on an inventory of 43 existing indicator frameworks for cities indicators.

Other related Standards developed by ISO:

- **"ISO/TR 37150:2014 Smart community infrastructures — Review of existing activities relevant to metrics":**

  This Technical Report provides a review of existing activities relevant to metrics for smart community infrastructures.

  Community infrastructures – energy, water, transportation, waste, information and communications technology (ICT), etc. – support the operations and activities of communities and have a significant impact on economic and social development. They are a means towards ensuring the delivery of goods and services that promote economic prosperity and growth, and contribute to the quality of life.

  The existing activities relevant to metrics for "smart" community infrastructures were reviewed in this Technical Report. This Technical Report covers metrics which is designed to help buyers to evaluate technical performances of community infrastructure products and services for procurement. These metrics additionally can be used in real-time monitoring for the operation of an existing community infrastructure.

  It is expected that this Technical Report will be useful to the following individuals/groups:

  - national and local governments;
  - regional organizations;
  - community planners;
  - developers;
  - community infrastructure operators (e.g in the field of energy, water, waste, transportation, ICT);

- community infrastructure vendors (e.g. constructors, engineering firms, system integrators or component manufacturers);
- non-governmental organizations (e.g.. consumer groups).

This Technical Report uses a model of the community functions in Table 1 and reviews activities relevant to metrics for community infrastructures.

Layers	Examples of functions
Community services	Education, healthcare, safety and security, tourism, etc.
Community facilities (Contribution)	Residences, commercial buildings, office buildings, factories, hospitals, schools, recreation facilities, etc.
Community infrastructures	Energy, water, transportation, waste, ICT, etc.
NOTE "Water" includes sewage and wastewater as well as drinking water.	

Table 5 Layers of a community

As shown in the above table :

- Functions of community infrastructures are fundamental to support the other two layers;
- Products and services of community infrastructures are more technology-oriented, more internationally-tradable than those in other layers and therefore appropriate for international standardization.

This Technical Report is intended to be used in the following ways:

- as a reference document
- to analyze the commonalities and gaps in existing activities relevant to metrics on smart community infrastructures
- to review the value of deploying smart community infrastructures
- as a basis for future standardization
- to assist stakeholders to have a better understanding of state-of-the-art smart community infrastructures around the world

(https://www.iso.org/obp/ui/#iso:std:iso:tr:37150:ed-1:v1:en)

- **"ISO/TS 37151:2015 Smart community infrastructures — Principles and requirements for performance metrics":**

  This standard was last reviewed and confirmed in 2018. Therefore this version remains current. ISO/TS 37151:2015 gives principles and specifies requirements for the

  - definition,

- identification,
- optimization, and
- harmonization

of community infrastructure performance metrics, and gives recommendations for analysis, including

- smartness,
- interoperability,
- synergy,
- resilience,
- safety, and
- security

of community infrastructures.

Community infrastructures include, but are not limited to, energy, water, transportation, waste, and ICT.

The principles and requirements of ISO/TS 37151:2015 are applicable to communities of any size sharing geographic areas that are planning, commissioning, managing, and assessing all or any element of its community infrastructures.

(https://www.iso.org/standard/61057.html)

3.5 Security Standards

NIST- The National Institute of Standards and Technology: NIST has just released Version 1.0 of the *NIST Privacy Framework: A Tool for Improving Privacy through Enterprise Risk Management* for managing privacy risk in order to help organizations keep this balance. Developed from a draft version in collaboration with a range of stakeholders, the framework provides a useful set of privacy protection strategies for organizations that wish to improve their approach to using and protecting personal data. The publication also provides clarification about privacy risk management concepts and the relationship between the Privacy Framework and NIST's Cybersecurity Framework. (https://www.nist.gov/news-events/news/2020/01/nist-releases-version-10-privacy-framework)

ISO/IEC 27032:2012 - (Guidelines for cybersecurity) provides guidance for improving the state of Cybersecurity, drawing out the unique aspects of that activity and its dependencies on other security domains, in particular:

- information security,
- network security,

- internet security, and
- critical information infrastructure protection (CIIP).

It covers the baseline security practices for stakeholders in the Cyberspace. This International Standard provides:

- an overview of Cybersecurity,
- an explanation of the relationship between Cybersecurity and other types of security,
- a definition of stakeholders and a description of their roles in Cybersecurity,
- guidance for addressing common Cybersecurity issues, and
- a framework to enable stakeholders to collaborate on resolving Cybersecurity issues.

ISO/IEC 27001:2005 - (Information security management) covers all types of organizations (e.g. commercial enterprises, government agencies, not-for profit organizations). ISO/IEC 27001:2005 specifies the requirements for establishing, implementing, operating, monitoring, reviewing, maintaining and improving a documented Information Security Management System within the context of the organization's overall business risks. It specifies requirements for the implementation of security controls customized to the needs of individual organizations or parts thereof. ISO/IEC 27001:2005 is designed to ensure the selection of adequate and proportionate security controls that protect information assets and give confidence to interested parties.

ISO/IEC 27001:2005 is intended to be suitable for several different types of use, including the following:

- use within organizations to formulate security requirements and objectives;
- use within organizations as a way to ensure that security risks are cost effectively managed;
- use within organizations to ensure compliance with laws and regulations;
- use within an organization as a process framework for the implementation and management of controls to ensure that the specific security objectives of an organization are met;
- definition of new information security management processes;
- identification and clarification of existing information security management processes;
- use by the management of organizations to determine the status of information security management activities;
- use by the internal and external auditors of organizations to determine the degree of compliance with the policies, directives and standards adopted by an organization;
- use by organizations to provide relevant information about information security policies, directives, standards and procedures to trading partners and other organizations with whom they interact for operational or commercial reasons;
- implementation of business-enabling information security;
- use by organizations to provide relevant information about information security to customers.

ISO/IEC 27005:2011 - (Information security risk management) provides guidelines for information security risk management. It supports the general concepts specified in ISO/IEC 27001 and is designed to assist the satisfactory implementation of information security based on a risk management approach. Knowledge of the concepts, models, processes and terminologies described in ISO/IEC 27001 and ISO/IEC 27002 is important for a complete understanding of ISO/IEC 27005:2011.

ISO/IEC 27005:2011 is applicable to all types of organizations (e.g. commercial enterprises, government agencies, non-profit organizations) which intend to manage risks that could compromise the organization's information security.

3.6 Other Related Standards

"ISO 37101:2016 Sustainable development in communities — Management system for sustainable development — Requirements with guidance for use"

This International Standard gives a holistic approach in order to the establishment of the requirements of a management system for sustainable development in communities, **including cities**. It provides guidance aimed at:

- improving the contribution of communities to sustainable development;
- fostering smartness and resilience in communities, while taking into account the territorial boundaries to which it applies;
- assessing the performance of communities in progressing towards sustainable development.

This International Standard sets out the requirements and the guidance to help communities achieve a framework that will allow them to become more sustainable.

This International Standard fosters the establishment of a multi-actor process in communities, through a holistic approach that facilitates the cooperation of all interested parties and avoids a silo approach. It is intended to provide guidance for organizations that implement other management systems that are compatible with this International Standard, such as ISO 14001, ISO 45001, ISO 50001, ISO 20121, ISO 14046 and ISO 26000, whether involved directly or indirectly in sustainable development in communities at different stages in their life cycles.

Involvement of interested parties through a multi-actor process can take different forms, such as:

- participatory partnerships;
- public participation;
- community based collaboration.

They all aim at involving interested parties in a cooperative dialogue for more sustainable solutions.

Successful implementation of this International Standard can:

- help to build consensus on sustainable development within communities;

- improve the sustainability, smartness and resilience of strategies, programmes, projects, plans and services conducted under the direct responsibility of communities, or on the territory they relate to;
- evolve cross-se

(https://www.iso.org/obp/ui/#iso:std:iso:37101:ed-1:v1:en)

ISO/TR 37152:2016 Smart community infrastructures — Common framework for development and operation:

ISO/TR 37152 outlines basic concept of a common framework for the development and operation of smart community infrastructures.

This framework;

- describes the planning, development, operation and maintenance methodology to facilitate the harmonization of each infrastructure as a part of a smart community and ensures that the interactions between multiple infrastructures are well orchestrated.
- be applicable to all processes of smart community infrastructures' life cycle (from conceptual design through planning, development, operation, maintenance, redevelopment and feedback). The infrastructures to be covered are energy, water, transportation, waste management, ICT and others.
- be adoptble by all relevant stakeholders who are engaged in planning, development and operation of smart community infrastructures, including planners, developers, business operators and suppliers.
- aims to cover the processes in which these stakeholders are engaged, such as management, organizational structure, analyses and design methods, and documentations.
- aims to ensure consistency between smart community infrastructures without overlapping with existing work (see Figure below). It incorporates the metrics as a KPI of the development, operation and maintenance methodology.

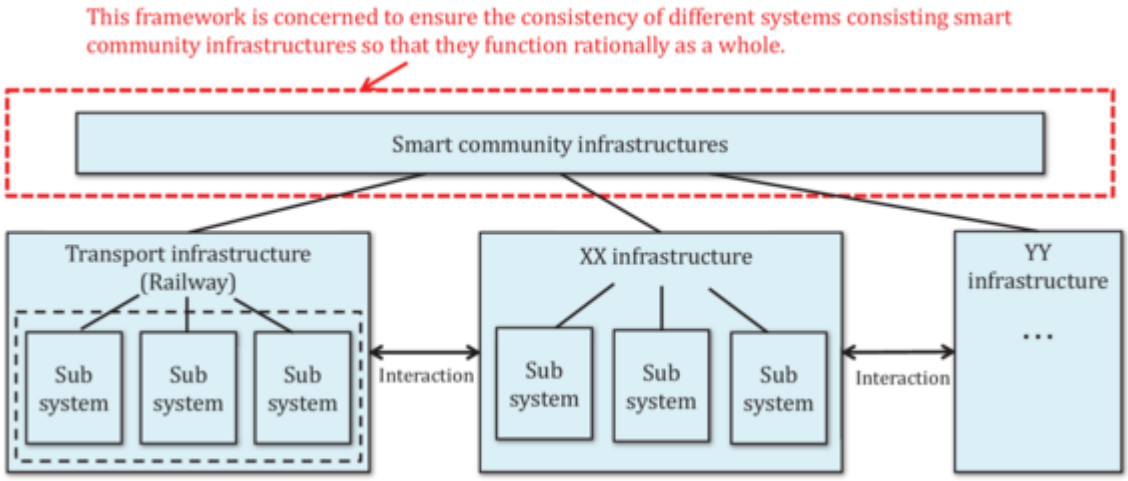

Figure 8 Scope of the framework

(https://www.iso.org/obp/ui/#iso:std:iso:tr:37152:ed-1:v1:en)

4. Needs Analysis of Cities

This section provides an in-depth analysis on the results of the University of Rotterdam survey. This study will allow general conclusions to be drawn on the key needs of the cities surveyed and will serve to have a global vision of the use of digital technologies in European and international cities.

We will then proceed with a point-by-point analysis of the results obtained, followed by a brief final conclusion at the end.

4.1 MARKET UPTAKE

According to the referenced study, a high percentage of cities are currently in development or implementation phases of various urban data platforms. These data show the fact that there is a certain interest by European cities about the collection and use of this information for the benefit of the entire society.

The adoption of this type of platforms is also growing considerably, especially in the last 5 years, where around 30% of the cities that are carrying out digitization activities and projects. This percentage is expected to rise to 50% by 2024.

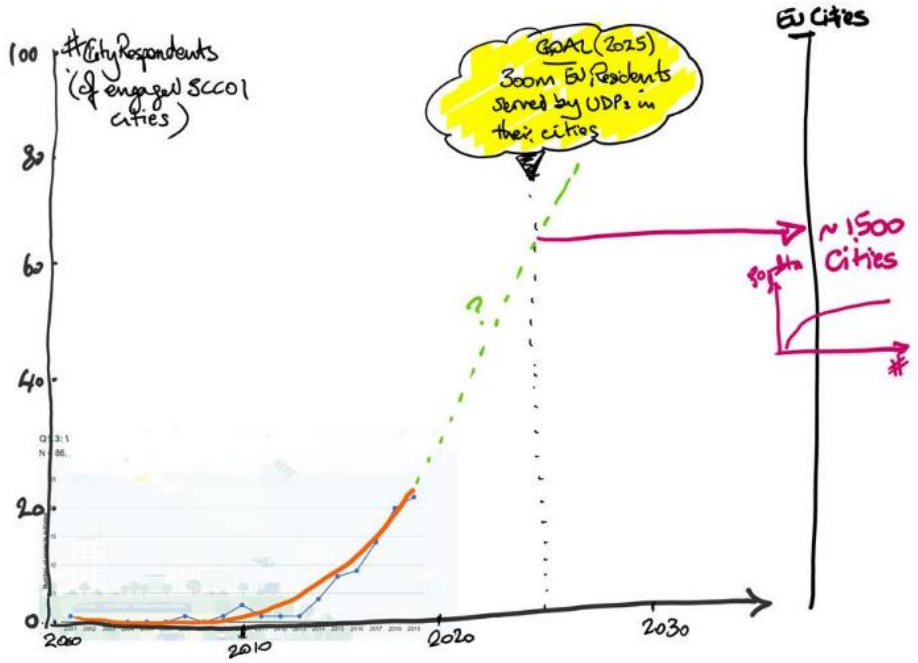

Figure 9 Adoption of UDP systems[74]

Regarding the systems used in these cases, 75% of the cities have 10 fully functioning systems or less. It should be noted that the real use of these systems by the administration, society and companies is lower than expected.

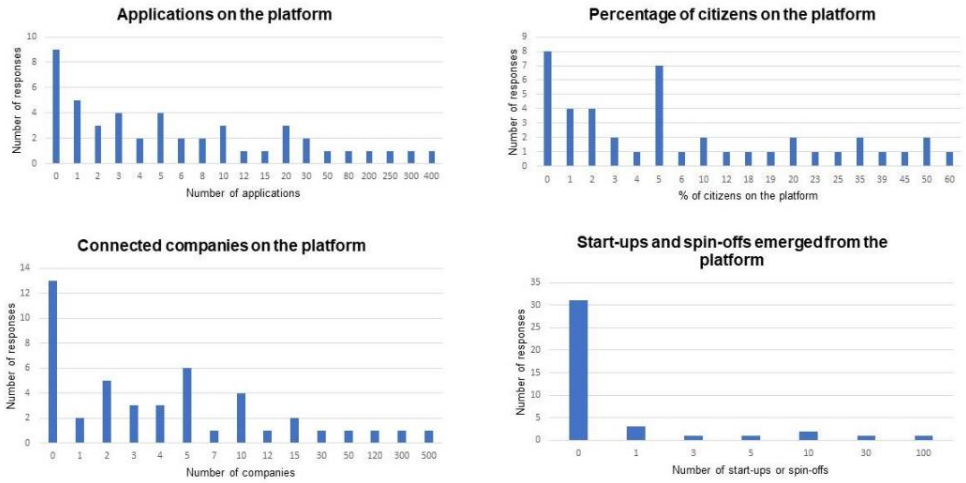

Figure 10 Numerous graphics based on survey's results[74]

In this field, the recommended thing to do would be to stimulate the use of new processes, throughout different programs offered by the European Union dedicated to this topic of smart cities. We must also study how to improve the percentage of use of existing systems. This could be done

through courses and other educational programs, aimed at teaching the functionalities of these platforms, on how they can improve the lives of their citizens and administrations.

4.2 PURPOSE & SCOPE OF URBAN DATA PLATFORMS

50% of the cities surveyed have clear ambitions to establish an open and interoperable data platform that encompasses the entire city and supports multiple systems. In this way, they want to try to achieve a certain independence grade of the platforms and systems in use, in case they want to change services in the future.

It is also identified that 66% of the existing platforms are publicly owned, and more than 80% are influenced by the city's public agencies. In general, it can be said that the city council of each city is the main responsible for the organization, implantation and use of these systems within each locality.

The key to improve these platforms within the city lies in a better communication between the body that manages these platforms, and the rulers of each city. In this way, the key objectives of each organization and city can be clarified, in order to cover each use case that may have the most impact within each locality. Thus, it is intended to develop a specific plan for each city, where the specific needs of each one is reflected.

4.3 STAKEHOLDER PARTICIPATION

It has been shown in this study that society is not generally interested in, or does not trust, the services provided by these platforms. To solve this, services that are making a difference in leading cities in this sector should be identified and considered implemented in more cases.

Another step to solve this situation would be to involve citizens to a greater extent within these platforms, offering them services that are easy to use, as well as useful. Finally, full control of the data by users could also be guaranteed, thus showing them a clear commitment to the privacy of the data generated.

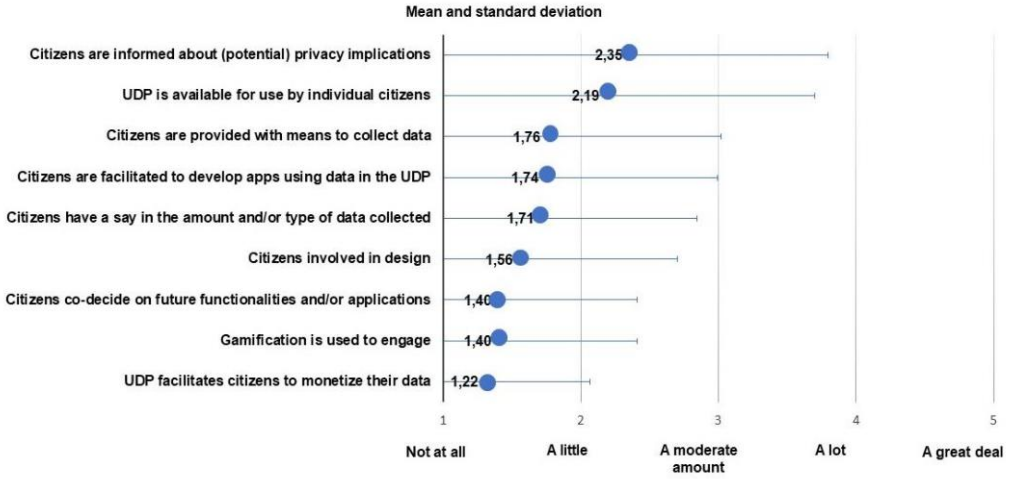

Figure 11 Society engagement infographic[74]

This links perfectly with another of the key points with which to improve the relationship between society and these platforms: trust. For these types of systems, in order to work properly and be fully integrated into the day-to-day lives of people, they must go unnoticed by their users. Thus, it is necessary to offer a complete and total transparency from these services to the citizens. More and more it is being seen that people are concerned about their online privacy, and they will not accept a system that does not offer them control over their private data.

Finally, the last thing to be analyzed here will be which role the legislation system of each city plays, regarding the implementation of these systems. According to the study, there are multiple bureaucratic barriers, mostly related to the security and privacy of information by all the agents involved. As a base point, all systems in use today, or in the future, must comply with the GDPR regulations of the European Union. The creation of applications is also recommended so that both users and administration can have full control and unlimited access to all their saved data. Similarly, the use of open and internationally agreed standards is also encouraged.

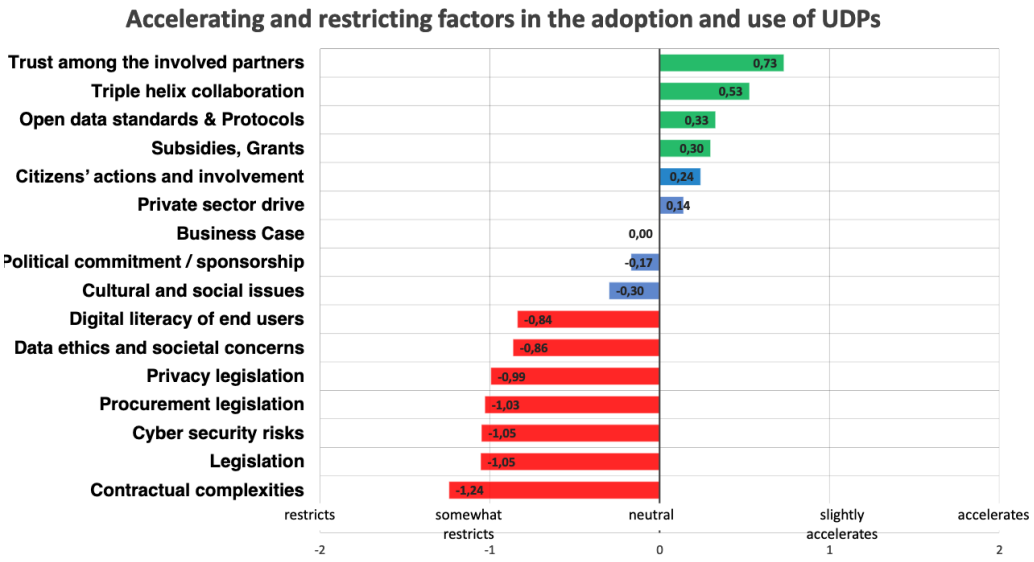

Figure 12 Accelerating and restricting factors in the adoption and use of UDPs[74]

4.4 CAPACITY BUILDING

This European study states that 42% of the surveyed cities have some kind of data managers and administrators within the city council itself. To improve their training and knowledge on the subject, it is proposed to create a network of this type of position at European level, with the aim of helping the exchange of information between localities.

It is also highlighted, as has been established in previous sections, the commitment to the implementation of open and free standards, which help collaboration between silos and data warehouses, located in different cities and regions. This can pleasantly help communication and collaboration between different administrations and countries.

Finally, from this study it can also be made clear that the free distribution of information among the different agents is very beneficial for the correct implementation and development of the different urban data platforms. More than 70% of the respondents share the information to their citizens through a free and free platform, while another 49% share the information through certain APIs to different service platforms and collaborating third-party companies.

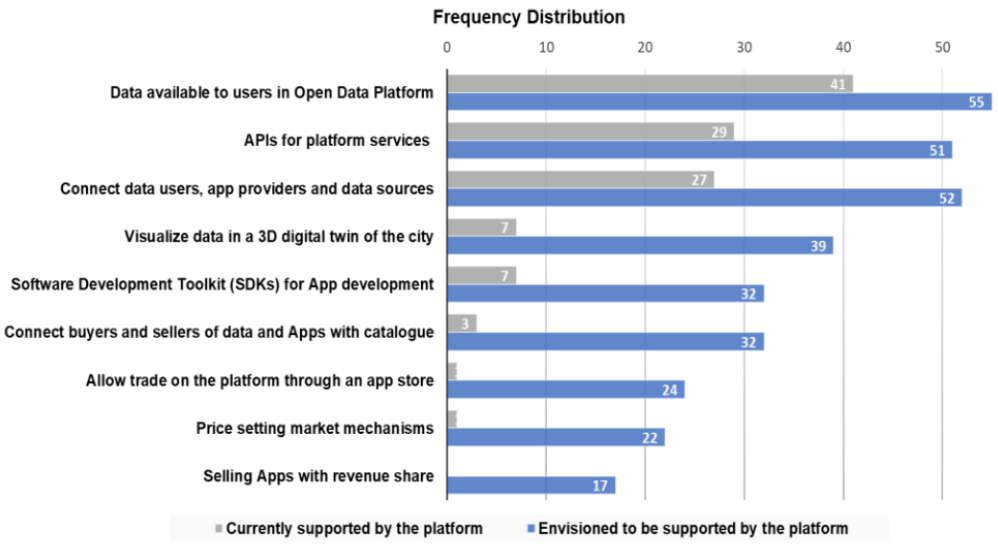

Figure 13 Data sharing infographic[74]

4.5 FINANCIAL MATTERS

Another fundamental issue regarding the implementation of these intelligent systems is the means of financing available to cities in general. The wide use of new technologies makes the implementation of these services more expensive.

The vast majority of cases studied here declare that there are clear problems related to the identification of clear use cases that open the doors for them in terms of financing channels. It is also established that more than 80% of the financing comes from the public sector, and that 60% are financed independently.

To solve this financing problem, the time and price of implementation of these services must be substantially improved, making them more attractive to potential investors. The robustness of the tools to be used must also be improved.

All these recommendations are intended to be developed in the medium and long term, but as technology advances and becomes cheaper, it should be possible to see an increase in the implementation of these platforms.

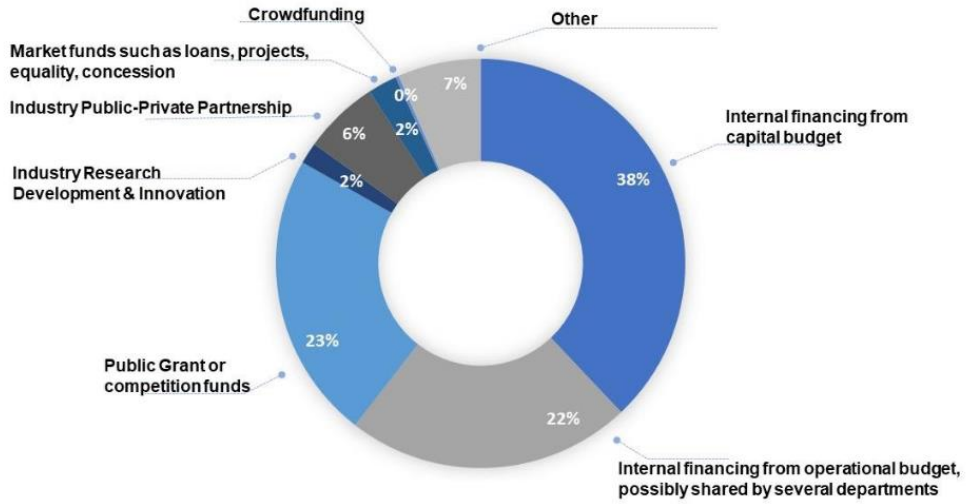

Figure 14 UDPs' budget & financial sources[74]

4.6 RISK & BARRIERS

Finally, the last point related to these studies is related to the financial risks linked to the implementation of these kind of platforms in the city. This last survey underlines what's previously said: the most important factor is to develop a real & sustainable set of use cases. Those will drive innovation forward, developing trust with the implementer of the platform and investors in the way.

The availability of clear use cases will drive the use of these systems from the target audience: the inhabitants of a city. If they're useful and clear to them they will simply use them, instead of fighting against systems that they don't understand.

Another key point relates to the avoidance of vendor lock-in at all cost. This will mean that the city has full independence about the decision of collaborating with other partners in the future, something that has been underlined already in previous sections of this study.

5. REFERENCES

[1] Millard, J.: Open Governance Systems: Doing More with More. Government Information Quarterly (2015).

[2] Janssen, M., & Helbig, N.: Innovating and changing the policy-cycle: Policy-makers be prepared! Government Information Quarterly (2015).

[3] Lasswell, H.D.: The policy orientation. The Policy Sciences. Stanford: University Press.

[4] Bertot, J. C., & Choi, H.: Big data and e-government: issues, policies, and recommendations. Proceedings of the 14th Annual International Conference on Digital Government Research, dg.o. 2013, June 17-20, Quebec City, Canada, pp. 1-10 (2013).

[5] Janssen, M., Charalabidis, Y. & Zuiderwijk, A.: Benefits, Adoption Barriers and Myths of Open Data and Open Government. Information Systems Management, vol. 29, no. 4, pp. 258-268 (2012).

[6] Ferro, E., Loukis, E.N., Charalabidis, Y. & Osella, M.: Policy making 2.0: From theory to practice, vol. 30, pp. 359-368 (2013).

[7] Linders, D.: From e-government to we-government: Defining a typology for citizen coproduction in the age of social media. Government Information Quarterly, vol. 29, pp. 446-454 (2012).

[8] Bijlsma, R.M., Bots, P.W.G., Wolters, H.A. & Hoekstra, A.Y.: An Empirical Analysis of Stakeholders' Influence on Policy Development. Ecology and Society, vol. 16, no.1, pp. 51-66.

[9] Davies, W.: How statistics lost their power – and why we should fear what comes next. The Guardian, 19 January (2017), https://www.theguardian.com/politics/2017/jan/19/crisis-of-statistics-big-data-democracy.

[10] Williamson, B.: Governing methods: policy innovation labs, design and data science in the digital governance of education. Journal of Educational Administration and History, vol. 47, no. 3, pp. 251-271 (2015).

[11] Fuller, M., & Lochard, A.: Public policy labs in European Union Member States; EUR 28044 EN June (2016), http://publications.jrc.ec.europa.eu/repository/bitstream/JRC102665/final%20report%20w%20identifiers.pdf.

[12] Gil-Garcia, J.R., Zhang, J. & Puron-Cid G.: Conceptualizing smartness in government: An integrative and multi-dimensional view. Government Information Quarterly, vol. 33, pp. 524-534 (2016).

[13] Janssen, M. & Estevez, E.: Lean government and platform-based governance – Doing more with less. Government Information Quarterly, vol. 30, pp. S1-8 (2013).

[14] Meijer, A.: E-governance innovation: barriers and strategies. Government Information Quarterly, vol. 32, pp. 198-206 (2015).

[15] Gil-Garcia, J.R. & Sayogo, D.S.: Government inter-organizational information sharing initiatives: Understanding the main determinants of success. Government Information Quarterly (2016).

[16] Hui, G. & Hayllar, M.R.: Creating Public Value in E-Government: A Public-Private-Citizen Collaboration Framework in Web 2.0. The Australian Journal of Public Administration, vol. 69, no. S1, pp. S120–S131 (2010).

[17] Koussouris, S., Lampathaki, F., Kokkinakos, P., Askounis, D. & Misuraca, G.: Accelerating Policy Making 2.0: Innovation directions and research perspectives as distilled from four standout cases. Government Information Quarterly, vol. 32, pp. 142-153 (2015).

[18] Moore, M.H.: Creating Public Value Strategic Management in Government. Harvard University Press (1995).

[19] Stoker, G.: Public Value Management A New Narrative for Governance? The American Review of Public Administration, vol. 36, no. 1, pp. 41-57 (2006).

[20] Bannister, F. & Connolly, R.: ICT, public values and transformative government: A frame-work and programme for research. Government Information Quarterly, vol. 31, pp. 119-128 (2014).

[21] Dawes, S.S.: The Evolution and Continuing Challenges of E-Governance. Public Admin-istration Review, vol. 68, no. 4, pp. S86-102 (2008).

[22] McDermott, P.: Building open government. Government Information Quarterly, vol. 27, pp. 401-413 (2010).

[23] Chadwick, A.: Bringing E-Democracy Back In Why it Matters for Future Research on E-Governance. Social Science Computer Review, vol. 21, no. 4, pp. 443-455 (2003).

[24] Schroeder, R. (2016). Big data business models: Challenges and opportunities. Cogent Social Sciences, 2(1), 1166924. http://doi.org/10.1080/23311886.2016.1166924

[25] Vierimaa & all: Final business and value networks. ITEA2 CAP project deliverable 5.1.2. (2016).

[26] URBAN AGENDA for the European Union, Partnership for Digital Transition. ORIENTATION PAPER.
https://ec.europa.eu/futurium/en/system/files/ged/digital_transition_orientation_paper.pdf

[27] Quality in open data based digital service ecosystem. Doctoral Thesis, Anne Immonen, 2017.. VTT Science 159.

[28] Immonen, A., Palviainen, M., & Ovaska, E. (2014). Requirements of an open data based business ecosystem. IEEE Access, 2, 88–103. doi:10.1109/ACCESS.2014.2302872

[31] Douglas W. Gage. Sensor abstractions to support many-robot systems. In Proceedings of SPIE Mobile Robots VII, pages 235–246, 1992.

[32] Gentili, M.; Mirchandani, P.B. Locating sensors on traffic networks: Models, challenges and research opportunities. Transp. Res. Part C Emerg. Technol. 2012, 24, 227–255.

[33] https://nos.nl/artikel/2198410-zo-weinig-geluidsoverlast-bij-lelystad-dat-kon-niet-kloppen-dacht-deze-man.html [in Dutch]

[34] https://www.samenmetenaanluchtkwaliteit.nl/projecten/meetcampagne-no2-milieudefensie [in Dutch]

[35] https://milieudefensie.nl/publicaties/rapporten/zelf-meten-is-zeker-weten [in Dutch]

[36] Jacobs, Piet & Richard Borsboom, Ir & , Kemp. (2016). PM2.5 in Dutch Dwellings due to Cooking.

[37] Hagens WI, van Overveld AJP, Fischer PH, Gerlofs-Nijland ME, Cassee FR, Gezondheidseffecten van houtrook : Een literatuurstudie. RIVM Rapport 609300027, 2012. [in Dutch]

[38] Creating Value through Open Data: Study on the Impact of Re-use of Public Data Resources

https://www.europeandataportal.eu/sites/default/files/edp_creating_value_through_open_data_0.pdf

[39] Berntzen L., Johannessen M.R., & Florea A. : Sensors and the Smart City. Fifth International Conference on Smart Cities, Systems, Devices and Technologies, Valencia, Spain, Volume: ISBN 978-1-61208-4763 (2016)

[40] Park T, Abuzainab N, Saad W. “Learning how to communicate in the Internet of Things: Finite resources and heterogeneity.” IEEE Access, 2016, vol. 4, pp. 7063-73.

[41] Kumar N, Chilamkurti N, Misra SC. “Bayesian coalition game for the Internet of things: an ambient intelligence-based evaluation.” IEEE Communications Magazine. 2015, 53(1), pp.48-55.

[42] Silva BN, Khan M, Han K. “Integration of Big Data analytics embedded smart city architecture with RESTful web of things for efficient service provision and energy management.” Future Generation Computer Systems. 2017.

[43] Piovesan N, Turi L, Toigo E, Martinez B, Rossi M. “Data analytics for smart parking applications.” Sensors, 2016,16(10), pp. 1575.

[44] Barnaghi P, Wang W, Henson C, Taylor K. “Semantics for the Internet of Things: early progress and back to the future.” International Journal on Semantic Web and Information Systems (IJSWIS), 2012, 8(1), pp. 1-21.

[45] Ganzha M, Paprzycki M, Pawłowski W, Szmeja P, Wasielewska K. “Semantic interoperability in the Internet of Things: An overview from the INTER-IoT perspective.” Journal of Network and Computer Applications, 2017, 81, pp. 111-124.

[46] Kovacs E, Bauer M, Kim J, Yun J, Le Gall F, Zhao M. “Standards-based worldwide semantic interoperability for IoT.” IEEE Communications Magazine. 2016, 54(12), pp. 40-46.

[47] Alaya MB, Medjiah S, Monteil T, Drira K. “Toward semantic interoperability in oneM2M architecture.” IEEE Communications Magazine, 2015, 53(12), pp. 35-41.

[48] Jabbar S, Ullah F, Khalid S, Khan M, Han K. “Semantic interoperability in heterogeneous IoT infrastructure for healthcare.” Wireless Communications and Mobile Computing, 2017.

[49] Gyrard A, Bonnet C, Boudaoud K, Serrano M. “Assisting IoT projects and developers in designing interoperable semantic web of things applications.” In IEEE International Conference on Data Science and Data Intensive Systems (DSDIS), 2015, pp. 659-666.

[50] Gyrard A, Serrano M. “Connected smart cities: Interoperability with SEG 3.0 for the Internet of things.” In 30th International Conference on Advanced Information Networking and Applications Workshops (WAINA), 2016, pp. 796-802.

[51] Agarwal R, Fernandez DG, Elsaleh T, Gyrard A, Lanza J, Sanchez L, Georgantas N, Issarny V. “Unified IoT ontology to enable interoperability and federation of testbeds.” In IEEE 3rd World Forum on Internet of Things (WF-IoT), 2016, pp. 70-75.

[52] Hodges J, García K, Ray S. “Semantic Development and Integration of Standards for Adoption and Interoperability.” Computer, 2017, 50(11), pp. 26-36.

[53] Claveria, Oscar, and Jordi Datzira. "Forecasting tourism demand using consumer expectations." Tourism Review 65.1 (2010): 18-36.

[54] Dritsakis, Nikolaos, and Spiros Athanasiadis. "An econometric model of tourist demand: The case of Greece." Journal of hospitality & leisure marketing 7.2 (2000): 39-49

[55] Palmer, Alfonso, Juan Jose Montano, and Albert Sesé. "Designing an artificial neural network for forecasting tourism time series." Tourism Management 27.5 (2006): 781-790

[56] Guizzardi, Andrea, and Annalisa Stacchini. "Real-time forecasting regional tourism with business sentiment surveys." Tourism Management 47 (2015): 213-223

[57] Les Rencontres de la Mobilité Intelligente 2016 (Smart Mobility encounters 2016), B. Geroudet, STERELA (2016)

[58] Improving Traffic Prediction with Tweet Semantics. IJCAI (2013)

[59] Social media based transportation research: the state of the work and the networking. IEEE/CAA Journal of Automatica Sinica. (2017).

[60] ITU-T, (2014). Smart Sustainable Cities: An Analysis of Definitions. Tech. Report. FG-SSC-10/2014, (2014).

[61] Opensource.com. (2018). Open source FIWARE platform creates new IoT business opportunities. [online] Available at: https://opensource.com/business/16/11/fiware-platform [Accessed 9 Feb. 2018].

[62] Santana, E., Chaves, A., Gerosa, M., Kon, F. and Milojicic, D. (2017). Software Platforms for Smart Cities. ACM Computing Surveys, 50(6), pp.1-37.

[63] Clout-project.eu. (2018). ClouT concept: The Cloud of Things | ClouT. [online] Available at: http://clout-project.eu/clout-concept-the-cloud-of-things/ [Accessed 9 Feb. 2018].

[64] Theregister.co.uk. (2018). Cambridge wheels out latest smart city platform, ready for devs. [online] Available at: https://www.theregister.co.uk/2017/03/23/cambridge_intelligent_city_platform_iot_launch/ [Accessed 9 Feb. 2018].

[65] Festival. (2018). FESTIVAL: Heterogeneous Testbed Federation Across Europe and Japan. [online] Available at: http://www.festival-project.eu/en/?page_id=1839 [Accessed 9 Feb. 2018].

[66] G. Moldovan, E. Z. Tragos, A. Fragkiadakis, H. C. Pohls, D. Calvo, "An IoT middleware for enhanced security and privacy: the RERUM approach", New Technologies Mobility and Security (NTMS) 2016 8th IFIP International Conference on, pp. 1-5, 2016, November.

[67] Bin Cheng, Salvatore Longo, Flavio Cirillo, Martin Bauer, and Ernoe Kovacs. 2015. Building a big data platform for smart cities: Experience and lessons from santander. In 2015 IEEE International Congress on Big Data (BigData Congress’15). IEEE, 592–599.

[68] ISO, (2018). Iso.org Retrieved 13 February 2018, from http://iso/IECAWI30145Informationtechnology-SmartcityICTreferenceframework

[69] ISO, (2018). Smart City Report, Retrieved 13 February 2018, from http://www.iso.org/iso/smart_cities_report-jtc1.pdf

[70] ISO 30145, (2018). ISO/IEC AWI 30145 - Information technology -- Smart city ICT reference framework. Iso.org. Retrieved 13 February 2018, from http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=70301

[71] IEEE SA - 2413 - Standard for an Architectural Framework for the Internet of Things (IoT). (2018). Standards.ieee.org. Retrieved 13 February 2018, from http://standards.ieee.org/develop/project/2413.html

[72] Smart City Framework. (2018). Pages.nist.gov. Retrieved 13 February 2018, from https://pages.nist.gov/smartcitiesarchitecture/community/

[73] Standard. (2018). Hypercat. Retrieved 13 February 2018, from http://www.hypercat.io/standard.html

[74] Insights from the EIP-SCC network: https://eu-smartcities.eu/sites/eu-smartcities.eu/files/2019-01/Connected%20Smart%20Cities%2017-1-2019%20-%20M%20van%20Oosterhout%20%28draft%29%20final.pdf

[75] http://www.datex2.eu/sites/www.datex2.eu/files/Datex_Brochure_2011.pdf

[76] Claudine Metral, Roland Billen, Anne-Françoise Cutting-Decelle, Marie Van

[77] Open Geospatial Consortium, OGC City Geography Markup Language (CityGML) Encoding Standard. Version 2.0.0, 2012.

[78] Lorenz B, Ohlebach H.J, Yang L, “Ontology of Transportation Networks”, REWERSE-DEL-2005-A1-D2, REWERSE project, Project deliverable, 2005.

[79] A. Barredo Arrieta *et al.*, “Explainable Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI,” *Inf. Fusion*, vol. 58, pp. 82–115, Jun. 2020, doi: 10.1016/j.inffus.2019.12.012.

[80] Dlodlo, N., Gcaba, O., Smith, A., Internet of Things Technologies in Smart Cities, IIMC International Information Management Corporation, (2016), ISBN: 978-1-905824-55-7.

[81] Lung, C., Buchman, A., Sabou, S., Smart city emergency situations management system based on sensors network, IEEE 24th International Symposium for Design and Technology in Electronic Packaging (SIITME), (2018).

[82] Chaouch, H., Bayraktar, A.S., Celal Çeken, Energy Management in Smart Buildings by Using M2M Communication, IEEE Xplore, (2019).

[83] Spandana, G., Shanmughasundram, R., Design and Development of Air Pollution Monitoring System for Smart Cities, Proceedings of the Second International Conference on Intelligent Computing and Control Systems (ICICCS), (2018), ISBN:978-1-5386-2842-3.

6. ANNEX - Data In Smart Tourism

1 TRADITIONAL DATA SOURCES

Traditionally, the collection of tourism-specific data has been the responsibility of national and regional statistical offices. Depending on the importance of tourism in each country, the information has been obtained in very different ways, resulting in incomplete information in many cases. One of the challenges related to collection is the transversality and complexity of the tourist activity, as it is linked to different industries. Due to this, and due to the underestimation of the economic impact of the tourism sector by governments, it is the case that responsibilities are not clearly defined when collecting tourism data.

At the beginning of the 90s, the first agreement on Recommendations on Tourism Statistics (RTS) was published to create a common frame of reference to unify the work on tourism. Despite the fact that this frame of reference includes indicators about the nature and generic statistics of the sector in a country, including supply information, it mainly offers information on demand, such as tourist arrivals, volume, as well as the economic impact or the contribution to GDP or employment by the sector.

In Spain, a variety of tourism indicators are produced based on both the TSA (Tourism Satellite Accounts) reference framework and the IRTS (International Recommendations for Tourism Statistics). This information is generated through internal studies from surveys, which are completed with external data from interested parties.

1.2 METHODOLOGY AND INFORMATION

Currently, most of the information is collected in the traditional way, through surveys or other data sources (tickets, hotel occupancy, etc.). Certain information is collected directly by public authorities, while other data is generated through ad-hoc surveys implemented by private companies. In general, the frequency of data collection is low, this added to the fact that the data does not have a spatial component; it translates into a data of a static nature. The presentation of results is usually in the form of annual reports, while the raw data is usually available on different web pages of public entities.

Tourism information could be classified into three categories:

• Information on offer: allow a view of the total number of establishments and their structural and economic characteristics. In Spain, the INE carries out hotel occupancy surveys, a monthly survey of all hotel establishments located in the national territory. The information on offer, therefore, has information on open hotel establishments, estimated rooms, estimated places and degrees of occupancy (among others).

• Information on demand: in order to characterize, from an economic point of view, the different types of tourists (national versus international), the type of goods and services demanded, etc. From the aforementioned hotel occupancy survey, information is also available on incoming travelers, number of overnight stays and average stay. In addition, the INE collects information regarding entry to / exit from Spain through Frontur (surveys of tourist movements at borders), for international

users. These surveys provide information on foreign visitors at the level of main destination, reason for the trip, duration, expense (among others) as well as the Familitur and Egatur surveys focused on collecting information on tourist travel and spending by residents of Spain.

• Intrarelated tables of supply and demand: allow understanding the contribution of tourism to the economy through the use of macro variables, such as GDP, production or employment.

Regarding the information on offer related to the establishments and their attributes (e.g. quantity, type, location of official accommodation establishments), this is made public by public administrations. The source of these data is usually the establishments themselves, which communicate them to the administrations, so that they publish them at least once a year. In the case of Madrid, Spain, this information is accessible both through the Open Data platform and through the INE website.

The INE's sources of information are listed below:

• Frontur:

The Tourist Movement at the Border Statistics is a continuous survey whose main objective is to provide monthly and annual estimates of the number of non-resident visitors in Spain who arrive in our country (tourists and hikers), as well as the main characteristics of the trips that they carry out (access route, destination, country of residence, reason, form of organization, etc.).

• Familitur:

The Resident Tourism Survey is a continuous survey whose main objective is to provide monthly, quarterly and annual estimates of the trips made by the population resident in Spain and their main characteristics (destination, duration, reason, accommodation, means of transport, expense, sociodemographic characteristics of travelers, etc.)

• Egatur:

The Tourist Expenditure Survey is a continuous survey whose main objective is to know the tourist expenditure of foreign visitors when they leave Spain.

In addition, the different autonomous communities have their own tourism competencies, which are in charge of generating tourist reports for their corresponding field of analysis based on data from INE surveys, as well as through their own surveys that allow enriching the data.

1.3 STRENGTH AND WEAKNESS ANALYSIS

The strengths and weaknesses offered by using traditional data are discussed below.

Strengths:

• Easier to obtain data related to the nature of the trip (type of accommodation, total expense, reason for the trip, etc.)

• The data obtained through the survey is very complete, so it is less necessary to merge data to complement information

Survey data is a very complete type of data that allows the characteristics of users and their trips to be defined more precisely and specifically, which otherwise could only be inferred through a combination of different data sources.

Weaknesses:

• Need for survey planning

• Lower sample size, due to the high cost of sampling, both in terms of resources and time.

• Lower sampling frequency

• Occasionally true or incorrect information due to people simplifying their answers

• Need for the user's willingness to answer

• Possibility of having a biased sample

• Lack of information on special events that may influence mobility (weekends, holiday periods, relevant events, etc.)

The biggest drawback of traditional data sources is that it is not possible to have updated data in a short period of time, making it difficult to analyze the impact of certain measures or specific policies.

2 NOVEL DATA SOURCES

In the last years is has been tried to advance in the use of new data sources for completing the already existing traditional data sources, in order to get more valuable information which may be adding value to the different sectors.

2.1 NEW OPPORTUNITIES

Considering the previous information, there are several interesting aspects that show the opportunities offered by the use of non-traditional data, such as massive data sources:

• Neither the reference frameworks, nor the indicators developed in previous years contemplate the geospatial component.

• With the exception of the indicators proposed in 2018 in the EEA (European Economic Area) frame of reference, the indicators do not include information related to the flow of tourists, nor with other types of indicators or dynamic data sources, such as the geographical or temporal distribution of tourists.

• Only in recent years has attention begun to be paid to the possibilities offered by the fusion of data from different data sources (both traditional and non-traditional) to improve the analysis of the dynamic nature of tourism.

• Most of the information used in tourism analysis is still collected using traditional methods, which tend to be more time consuming, in addition to offering spatial granularity and low continuity, compared to new data sources

2.2 EXAMPLES OF APPLICATION TO THE TOURISM SECTOR

When talking about the use of big data in tourism, it is important to understand that there is a wide variety of data that is included under the term big data that, depending on its nature, can be used to contrast, complement or even complete other data sources. It is studied which data sources show the potential to complement or even substitute, with the correct fusion of different data sources, the traditional data sources.

The following are the most frequently used types of data:

• Point of sale (POS) transaction data: credit card usage data that allows characterizing the movements of tourists, as well as the economic impact

• image-based data: extracted from platforms such as Intagram, Flickr, etc. that allow to identify the main attractions of the destinations

• news and review data: obtained from platforms such as Twitter or Tripadvisor in order to obtain a better perception of the experiences and satisfaction of visitors

• Search data: Internet search data allows obtaining parameters for estimating tourist arrivals.

In addition to these new data sources, information about, for example, the accommodation offer or unofficial information about rates and room occupancy is increasingly being extracted through the use of third-party data platforms such as InsideAirbnb, open platforms data, as well as through search activities on the web, for example, online travel agency platforms and other metasearch engines such as Trivago, Expedia, etc.

Transaction data:

Transaction data is an important source of big data information for tourism, since it collects data on the use of credit cards, generating information for part of the activities that the user is carrying out in the destination, locating the same space temporarily, and on the other hand, the data that betrays the intention to travel, such as hotel payments, flights or tourist packages through the internet.

This type of data has a very high spatio-temporal definition, since it includes the exact time and place of the transaction. However, the frequency of generation of this data is restricted to the use that is given to the card, so it is not so useful when defining characteristics that require constant updating of the data, such as the flow between two points. This type of data mainly offers indicators of a specific trip, such as total expenditure or activities carried out at the destination.

Below are two examples of projects that have used this type of indicator and the data they have used for it:

New Zealand Ministry of Trade, Innovation and Employment

Between 2012 and 2013, New Zealand's Ministry of Commerce, Innovation and Employment (MBIE) launched two new regional tourism data collection methods using card transaction data: regional tourism indicators (RTI) and regional tourism estimators (RTE). These provide a detailed, specific and region-segmented view of the spending behaviors of national and international tourists. The first does so through monthly card expense reports, and the second with annual estimates of absolute spending. Thanks to this data, New Zealand regions are able to make more informed planning and investment decisions in tourism.

RTIs use aggregate electronic card transaction data classified by period, cardholder origin, merchant location, and merchant industry. Data is published as transaction accounts or expense ratios using any combination of these dimensions. Beyond their value for monitoring long-term growth and change in tourism spending, RTIs are also valued as a tool to assess the impact of events, marketing, and other values that influence tourism spending.

For their part, RTEs use card transaction data through a combination of data sources and modern statistical techniques to generate estimates of absolute expenditures that can be broken down by year, territorial authority, industry, and by country / region of origin of the cards. tourists. The method uses an iterative proportional adjustment to adjust total tourism spending from card transaction data to match, on the one hand, the totals by industry from the New Zealand Tourism Statistical Satellite Account, and another, with the totals by country of origin from the MBIE Survey on International Visitors. The regional results have been contrasted with several different regional indicators and the MBIE is confident that the estimates are reliable.

BBVA

In July and August 2014, Banco Bilbao Vizcaya Argentaria (BBVA) generated a report in which it analyzed more than 5.4 million anonymized and irreversibly aggregated transactions to obtain a vision of the spending habits of tourists. These data come from a sample of more than 2 million foreigners who used BBVA's POS to pay their expenses in Spain. The results are presented on an interactive map developed by Vizzuality. This report divides the data provided into expenses and transactions, which, in turn, are segmented according to the tourist's country of origin, the city / autonomous community in which the card is used, the recipient of the transaction and their sector. , and the date and time the expense occurs. Thanks to this segmentation, it is possible to evaluate which Spanish destinations tourists of different nationalities prefer, in which products and sectors they are willing to spend more and at what hours they are most active. In turn, these data contribute to determining the general contribution to the local economy by tourist spending –the amount spent by foreigners compared to that spent by Spaniards–, as well as better planning and estimation by the tourism sector.

Photographic data:

In recent years, with the emergence of Internet communities that share photos and social networks, many tourists publish their vacation images. Generally, information about where and when the images were taken is attached to these shared photos, which provides clues to establish the temporal and spatial distribution of the authors. Online photos generate data of great value for the

tourism field since it has information about the user, time of creation of the photo, time of upload to the platform, geographic information of the photo and text related to the photo. From this, it is possible to generate indicators related to tourism such as the places of greatest tourist interest, origin and destination of the trip or duration of the trip. Below are two examples of projects that have used this type of indicator and the data they have used for it:

Beijing Institute of Urban Environment, China

According to the Beijing Institute of Urban Environment, it is important to improve the understanding of the temporal and spatial behavior of tourists in order to improve the sustainable management of tourism. In their report (Li, Chunming & Zhao, Yang & Sun, Xiaoze & Su, Xiaodan & Zheng, Shuanning & Dong, Rencai & Shi, Longyu; 2011), they describe an approach based on free access photographs shared on the internet to analyze the temporal-spatial behavior of Chinese tourists identified from 1308 photographs taken in the Old Town of Lijiang. For this, the temporal variation of tourists, the length of stay, the daily average number of tourists, the traces of individual movements and the tourist attraction points are estimated. The results suggest that photo-based methods have obvious advantages. It is believed, in fact, that this wealth of publicly available data sources with spatio-temporal information will provide a new perspective in the analysis of the distribution of tourist flow and tourist attraction points and will be useful for planning tourism resources. These data are public, as well as the full report.

Data Science Laboratory at Warwick School of Business and Alan Turing Institute

A team of scientists from the Warwick Business School Data Science Laboratory and the Alan Turing Institute analyzed data from 69 million photos publicly shared on the Flickr platform between 2013 and 2014 to estimate countries' global tourism statistics. G7: Germany, Canada, the United States, France, Italy, Japan and the United Kingdom.

While countries often rely on time-consuming surveys at airports and accommodation to determine where their visitors are coming from, the new results show that inexpensive and readily available data from photos uploaded could provide similar measurements. .

In their report, they analyzed data on where and when those 69 million Flickr photos had been taken over a two-year period. To find out where the UK visitors were from, photos that had been taken in the UK were identified, and then examined where the photographer had taken photos in the last 12 months.

In order to carry out the study, some simplifying assumptions had to be made, such as that, between one photograph and the next, the authors remained in the country in which the last image was captured. However, even with this vast simplification, it was found that estimates of the number of travelers from different countries generated from online photo data were correlated with official tourism statistics, as in other G7 countries.

The G7 countries currently use various methods to calculate the number of tourists, including collecting data on airports, hotels and other tourist accommodation. However, all these methods have in common a delay in publication, which ranges from months to years. By contrast, using the near-instantaneous availability of online data could lead to much faster estimates of tourist flows between countries. By analyzing 69.2 million photos uploaded to Flickr in 2013 and 2014, the researchers were able to infer the travel patterns of nearly half a million people.

When these data were compared to official published by the G7, they found a strong correlation between estimates based on photos and official estimates.

News and Review Data:

Opinion data (such as opinions published on services such as TripAdvisor) have information not only on the places visited by users, but also on their evaluation. In this way it is possible to establish points of greater interest for tourists, as well as indicators of customer satisfaction.

This opinion data together with the characteristics of the users have been used by Eleonora Pantano, Constantinos-VasiliosPriporas, NikolaosStylos (2017) to predict the places with the greatest tourist attraction for users based on their characteristics.

To do this, the reviews published on tripadvisor for certain specific tourist attractions were analyzed. In a first step, those users whose reviews were very good (5/5) or very bad (0/5) were selected to differentiate between users who liked the point of interest and those who did not. These users were then characterized based on the tastes that they indicated in their profile. From this information, half of the sample was used to train a random forest model that predicted the result of the criticism based on the user profile, and the second half was validated with the model. The results showed the effectiveness of the model used.

Mobile phone data:

The telephony data collects the information generated by mobile devices when they interact with the telephone towers. The data has the following identifiers, location of the accessed telephony tower, time of telephone registration and nature of the registration (incoming / outgoing call, message, data packet ...). Given the great spatio-temporal granularity of this data, it is possible to obtain indicators of great value for tourism such as the OD flows of tourists, spatio-temporal distribution, areas with greater activity, among others.

There is an upward trend in the use of telephone data to obtain tourism indicators, mainly based on user mobility patterns. This type of data has gone from being a secondary data source, used to complement other data sources, to being a data source with the potential to replace traditional data sources.

The following are certain projects for which the use of mobile phone data in tourism has already been integrated:

In a study carried out by Ahas, R., Aasa, A., Mark, Ü., Pae, T., Kull, A. in 2007 the potential of the data to identify the seasonality of tourism in Estonia was analyzed. Seasonality was compared based on the volume of calls made by foreign users from the Estonian telephone network with the hotel

occupancy data, verifying that the patterns coincided in profile. On the other hand, the data was applied to detect the areas of greatest interest depending on the season from the variation in the volume of calls from foreigners in specific events, finding a variation in the areas of interest depending on the season of the year Finally, it is proposed to use this data to identify as yet unknown travel flows in the study area, such as Latvian fishermen flows to Lake Peipsi through the Tartu region.

On the other hand, Rein Ahas, Anto Aasa, Siiri Silm, Margus Tirub (2007) studied the distribution of foreigners in different study areas according to their nationalities. In this case, the profile of records was also compared with the distribution of hotel stays, resulting in very similar distributions. In this case, the locations visited by those users who had previously visited the Tartu area were studied, resulting in maps of tourist flows. The mobile phone data was also used in order to identify events that contributed to the increase in tourists in specific areas and periods. As expected, these peaks were detected that coincided temporally and geographically with the events.

Search data:

Internet search data (eg Google Trends), both in traditional search engines and metasearch engines, offer very important data for tourism. Today it is common to obtain most tourist services through the Internet, as well as information about destinations. These searches generate search volume data, which, when used correctly, can help predict different indicators of a tourist nature, such as tourist arrivals or places of greatest interest. According to existing studies, the results obtained through the use of this type of data have shown excellent performance, especially to capture the online behaviour of tourists and related decisions (Li et al., 2016).

2.3 STRENGTH AND WEAKNESS ANALYSIS

The use of mobile phone data for tourism has certain associated advantages and disadvantages. Among the advantages are the large size of the sample, since most people have a mobile phone and the frequency of generating records is high, the possibility of covering a larger spatial area, including areas with reduced numbers of visits, in addition to the additional information that this data hides, such as the user's nationality.

The main disadvantage associated with the use of mobile phone data resides mainly in the difficulty in accessing this type of data, which has slowed down the progress of studies related to this technology. This type of data shows weaknesses in the study areas where the density of telephone towers is low (or non-existent). The location of the user is restricted to the presence of these towers,

so in areas with low coverage, the number of records may be limited, in addition to the fact that the location of the user may differ greatly from the location of the tower.

Another of the difficulties that the use of this type of data presents is the expansion of the sample to the total population. While for residents in Spain the expansion can be carried out based on population data by census sections, foreign users who connect to the telephone network through roaming have to be expanded with data on the entrances and exits of tourists, which requires complex characterization processes for foreign users.

For this reason, national visitor tourism is more easily representable than international visitor tourism, which is usually represented in percentage terms to avoid errors derived from the sample expansion process.

3 COMPARATIVE TABLE

Indicator	Description	Traditional data source	Novel data source	Comparative
Tourism Satisfaction / Opinion	This indicator measures the degree of satisfaction of tourists visiting a tourist destination. This indicator is of interest to destinations because a high degree of satisfaction generates affinity of tourists to destinations, resulting in tourists who repeat visits.	Through surveys in tourist destinations it is possible to measure the degree of tourist satisfaction. These surveys measure user satisfaction on e.g. gastronomy, accommodation, transport, hygiene or safety issues.	Through opinion data sources, such as TripAdvisor or Twitter, it is possible to measure the degree of satisfaction with your hotels, restaurants or points of tourist interest. Those platforms that include scores allow more direct measurements than opinion platforms such as Twitter for which keyword analysis is required for the evaluation of messages, also known as sentiment analysis.	The results that can be obtained from the surveys allow a more static analysis of the quality of tourist services, given the size of the sample to be analyzed, given the nature of the data to be analyzed. However, the analysis of opinion data sources allows us to analyze whether, for example, a low quality in transport services is due to a specific event or if it is a phenomenon that is repeated over time.

Analysis of the behavior of tourism	This indicator is understood as the characterization of the activities carried out by the tourist in the tourist destination.	Surveys allow you to study the activity patterns carried out by the surveyed users.	The transaction data sources, POS, allow to know the expenditure made by users in a certain destination, each transaction being recorded, as well as the concept of these transactions.	Transaction data allows users to trace the spending patterns of users in a more disaggregated way, obtaining results such as e.g. Favorite time to eat or favorite places to shop. However, they do not provide information on total expenses and can only provide a snapshot of certain types of expenses.
Visited points of interest	This indicator makes it possible to measure the number of visits made to points of interest by visitors in a study area.	The administrations in charge of the points of interest are those that carry out the registration of volumes of visits received.	Visits made to points of interest can be estimated through the variation in the volume of opinions made by users on opinion platforms TripAdvisor or Twitter. In the same way, it would be possible through transaction data sources to record the transactions carried out at the specific point of interest.	The count of visitors to points of interest is more difficult to quantify with big-data data sources, because not all visits to points of interest record an opinion / transaction.
Volume of visits	The volume of visits measures the total number of visitors in a study area in a specific time interval.	Through surveys of both residents in Spain and foreigners, it is possible to estimate the volumes of visitors for specific areas at specific times.	Through big data data sources, such as mobile phone data, it is possible to capture people's visits, through a longitudinal analysis of users.	Mobile telephony data allow a greater disaggregation of the data, since surveys tend to measure monthly visitors, while mobile telephony data allow obtaining this same data on a daily or even hourly basis.
Overnight stays	The volume of overnight stays allows to	Traditional data sources include hotel occupancy	Using mobile phone data, it is possible to measure the number of	Big data data sources, such as mobile phone data, make it possible to

	measure the number of people staying overnight in an area.	data and visitor surveys to measure the number of visitors staying overnight in the study areas.	people staying overnight in the study areas from the analysis of the records generated by users during the night.	capture volumes of overnight stays that are not contemplated in traditional data sources, such as visitors staying overnight in non-hotel establishments such as e.g. Airbnb. They also allow a greater disaggregation of the data such as e.g. in daily overnight stays, instead of monthly or by nationality of the users.
Tourism demand prediction	Tourism demand predictions allow administrations to make improvements in resource management.	Through historical data and through time series analysis it is possible to make estimates of tourism demand.	Through the analysis of trends in social networks, such as Instagram or Twitter, it is possible to obtain indicators for tourist destinations such as travel intention or safety, which allow the creation of tourism demand prediction models.	Big data data sources offer the possibility of calculating indicators that will affect the forecast of demand in an updated way, introducing new dynamic variables to the model.