

Connecting Digital Cities: Return of Experience on the Development of a Data Platform for Multimodal Journey Planning

Jonathan Ouoba^(✉), Janne Lahti, and Jukka Ahola

VTT Technical Research Centre of Finland, Espoo, Finland
jonathan.ouoba@gmail.com, {janne.lahti,jukka.ahola}@vtt.fi

Abstract. The multiplication of real-time data sources in urban areas creates a fragmented environment regarding the supply of mobile services for transportation. Thus, in order to support the deployment of smart mobility services with a more global view of the urban context, the available data must be appropriately handled through dedicated platforms. It is in this context that the Connecting Digital Cities (CDC) project has been initiated at European level. The goal of the project is to deploy a platform that is able to collect, analyze and enrich the relevant information in order to provide a valuable aggregate of real-time data related to urban mobility. This paper presents a first return of experience in the development of this platform. We first provide an overview of the CDC project and we detail the architecture of the platform intended to collect and process real-time data. For validation purposes, we then describe a mobile service for multimodal journey planning that has been deployed to interact with the CDC platform. Finally, we share the lessons learned in terms of reliability of the system and proper integration of the system into urban ecosystems.

Keywords: Digital cities · Smart transportation · Journey planner · Mobile services · Mobility data

1 Introduction

The trend of continuous urbanization and the expansion of cities raise big challenges in the domain of transportation [6]. At a global scale, one million people move to cities every week and it is predicted that urban areas will be home to five billion people in 2020 [15]. This highlights the necessity for ICT-based solutions in order to rationalize the organization of urban spaces and also address issues related to pollution and energy preservation. At European level, many urban mobility services exist to improve the daily life of citizens in the domain of transportation. These solutions take advantage of real-time data provided by different actors of the urban environment. It concerns data from cities (e.g. traffic monitoring), public transportation and private operators (e.g. taxi and parking) that are made available to third parties through open APIs. However, the provided solutions give a fragmented view of urban areas and they are not always

complementary to each other. Indeed, due to the multiplication of data sources, they often lack a global view regarding urban mobility data. In addition, new types of information such as crowd-sourced data (accessible for instance via social networks) are too rarely considered.

In this context, EIT ICT Labs¹ is encouraging collaborative initiatives such as the Connecting Digital Cities (CDC) project that are intended to build interoperable solutions to support smart mobility services. This paper presents a first return of experience concerning the CDC project by discussing the following question: **How to appropriately combine real-time data from various sources, hence making it possible to support the development of mobile services for multimodal smart mobility?**

Our approach has been to build a platform with a relevant model from a technical and a business perspective. The remainder of the paper is organized as follows. First, we provide an overview of the CDC project and we detail the architecture of the platform² intended to collect and process real-time data in urban environments. Thereafter, we describe the Park&Ride service that has been deployed and evaluated in order to validate the functionalities of the CDC platform. Finally, we share the lessons learned in terms of reliability of the system and business model before concluding.

2 Connecting Digital Cities

2.1 Presentation of the Project

The Connecting Digital Cities project is carried out within the framework of EIT ICT Labs in the category of Urban Life and Mobility. The goal of the project is to deploy a platform that is able to collect, analyze and enrich the relevant information in order to supply a consistent aggregate of real-time data related to urban mobility. The objective is also to develop a set of mobile services making use of the platform so that its efficiency can be illustrated. In others words, this platform is intended to provide a relevant technical layer for supporting the development of transportation services for urban areas. Indeed, the proliferation of data sources makes it difficult to apprehend the whole reality of urban spaces and it leads to the deployment of fragmented mobility services. It concerns data from users (e.g. mobility traces, social networks), data describing the environment (e.g. weather conditions) as well as data produced by public authorities (e.g. public transportation, traffic flow). For interoperability and consistency purposes, it is thus essential to improve the framework to handle (in terms of collection and sharing) multiple sources of real-time information in an urban context. To this end, the straightforward approach consists in building a platform that helps to unify the vision of developers of services regarding the

¹ European Institute of Innovation and Technology - <http://www.eitictlabs.eu/>.

² For confidentiality reasons related to the conduct of the CDC project, all technical details are not provided. However, this limitation does not affect the contribution of the paper in regards with the return of experience.

available (real-time) data and the way to use it. This situation illustrates the relevance of the CDC project. The relevance is also reinforced by the actions promoted by EIT ICT Labs in the domain of Urban Life and Mobility [5].

The first phase of the CDC project took place between January and December in 2014. The consortium consists of industrial partners with big players in the domain of ICT-based technology such as Nokia and Thales, research institutes and SMEs offering mobile services for transportation. Three countries were involved, namely France, Finland and Italy. In terms of basis for the work, the project built on top of the results that have been achieved by former initiatives. These initiatives are:

- The ASSISTANT project [2,14] (an European initiative) the goal of which is to improve the mobility of elderly people by providing mobile services allowing them to safely and independently travel with public transportation;
- The SUS project [4,9] (an European initiative) which aims at developing interoperable technology bricks so that it enables cities to easily introduce advanced mobile services in the domain of transportation and tourism [7,8,13];
- The Tivit Digital Services framework which is part of a national project in Finland that aims at supporting the creation of user-driven digital services with the appropriate technological platforms [12].

In terms of architecture and functional characteristics, the next section discusses the approach for designing the CDC platform.

2.2 Architecture of the CDC Platform

As a reminder, the CDC platform is intended to provide a lightweight and service-oriented environment to support the effective development of urban mobility solutions. As such, the main concerns lie in efficiency and scalability. To handle these issues, the platform complies with the reference architecture for Urban Life and Mobility that is defined in the EIT ICT Action Line [5]. The access to the platform is based on the SOA [10] paradigm in order to foster interoperability. It is to be noted that this approach adheres to the principles of the European Interoperability Framework and the European Interoperability Reference Architecture [3].

In a concrete way, the architecture of the CDC system consists of three main layers (as illustrated Fig. 1) that can be described as follows:

- **Data Layer.** This module gathers and organizes the data from external sources. It enables the platform to regularly access remote servers (via HTTP-protocol) in order to collect data (in XML/JSON format) that is published by public/private stakeholders of the ecosystem related to urban mobility. It also ensures that the data is normalized (i.e. it makes use of the norm ISO 8601 [1] as the baseline for timestamps) so that it can be sorted, compared to other information and used by the other layers. Obviously, the pace of updating the data depends on the schedule of the providers. In our context, the Data Layer is able to handle data regarding weather conditions, traffic flows, parking facilities, taxi operators, public transportation and social networks.

- **Analytics Layer.** This module makes use of data collected through the Data Layer in order to discover patterns of urban mobility. It enables the platform to identify eventual traffic disruptions by analyzing the traffic flows, the state of the public transportation network and the relevant posts within social networks. Based on the extrapolation of data history and the anticipation of upcoming events, it also performs forecasts regarding the evolution of traffic flows and the availability of means of transport. Additional information is provided in the Subsect. 3.2 that describes the interaction of the platform with the server of a mobile service.

The merging of information from multiple sources improves the accuracy of traffic analysis and forecasts as it allows the discovery of significant events that can be “hidden”. To this end, this layer is endowed with spatiotemporal Complex Event Processing (CEP) capabilities so that it can perform deductions, analysis and correlations from elementary pieces of information.

- **Interface Layer.** This module provides public APIs so that third parties (e.g. developers, service providers) can take advantage of the analytics capabilities of the CDC platform. The APIs allow third parties to transmit HTTP-based requests specifying an area or a location. The Interface Layer processes these requests and interacts with the Analytics Layer in order to retrieve relevant information. Thus, third parties are able to receive real-time status and forecasts related to urban mobility (in terms of traffic conditions, public transportation) in a given area. This functionality is especially useful for mobile applications dedicated to urban mobility. As they usually rely on back-end systems to meet the expectations of users, it eases the process by merging the source of information.

In addition to these modules, a Registry is integrated to the system³. This component provides a comprehensive view of the capabilities of the CDC platform so that it can be effectively used by third parties. Through a web-based environment, it lists the available resources as well as the technical documentation associated to these resources. For evaluation purpose, the Registry also supports feedbacks of users (third parties) regarding the access to the platform.

2.3 CDC Ecosystem

From a business perspective, the potential of the CDC platform is emphasized by the fact that it eases the development of mobility services and it enhances the accuracy of information provided to end-users. Indeed, the capabilities of the platform (data processing and data mining) associated to the provision of public APIs give a simple and customized access to urban mobility data. Then, the service providers (developers of mobility services) can specialize and focus on core business by marketing their own services. We may also mention the public authorities in the list of important actors. Indeed, the public sector can consider the CDC system as an additional tool for planning, collecting and managing mobility issues regarding citizens.

³ This specific module has been developed in the framework of the E015 Digital Ecosystem project [16].

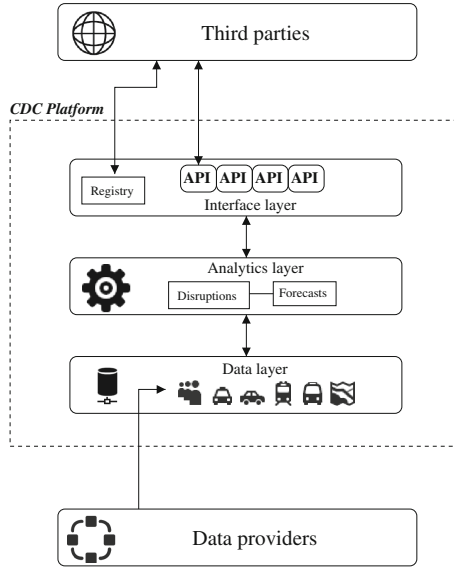


Fig. 1. Architecture of the CDC system

Other business aspects are provided in the Subsect. 3.3 related to the business model of the mobile service (Park&Ride) that has been developed within the CDC project.

3 Multimodal Journey Planner

3.1 Objective and Challenges

In order to illustrate the use of the CDC platform with a concrete case, a service was developed. This service is called Park&Ride and it consists of a mobile application interacting with a backend system to retrieve mobility data. Park&Ride aims at providing an appropriate tool to handle (real-time) multimodal journey planning in specific situations. The main usage scenario is as follows:

- First, the user chooses a destination address that he wants to reach;
- then, the application provides a navigation assistance making it possible to use personal transportation (e.g. private car) and/or monitor the route for traffic disruptions to propose alternative options with public transportation;
- finally, in case the user prefers an alternative solution because of a disruption, the application provides relevant information regarding the parking options and the adequate combination of public transportation to reach the final destination (with a mention of the corresponding walking distance to access public transportation facilities).

In line with the overall objectives of the CDC project, the Park&Ride service must specifically match the following requirements:

- **Quality of information and service.** It represents the added value for the user in terms of relevance of the information.
- **User friendliness.** It is related to the user experience when he interacts with the application via the dedicated interface.
- **Quality in the design of the system.** It concerns the security and the reliability of the system. The design must ensure that the system is resilient to faults while handling connectivity issues with smoothness and responsiveness.

The requirements that we have chosen to consider are returned to in the Subsect. 3.4 dedicated to the evaluation of the Park&Ride application.

3.2 Technical Architecture

Figure 2 illustrates the architecture of the Park&Ride service as well as its links with the CDC platform. As mentioned in the previous section, the service consists of two main entities, namely the Mobile Phone Application and the Park&Ride Server.

The Mobile Phone Application interacts with the Park&Ride Server by forwarding the requests of the user, by collecting the results of the requests and by retrieving relevant data regarding eventual disruptions in the traffic. Upon reception of data from the server, the application notifies the user by displaying the information and it prompts him to take an action (validation of a routing proposal, choice of an alternative solution in case of disruption, etc.). In addition, the Mobile Phone Application contains a map module (maintained by a third party) allowing it to route the user to the destination provided that he wishes to travel either via a private vehicle or via public transportation.

Regarding the Park&Ride Server, it is connected to the CDC platform and it acts as an intermediary for the Mobile Phone Application. The collected data, particularly with regard to the availability of parking lots and the detection of disruption, enables the server to determine the adequate combination (best parking options, transit via public transportation, routing with private vehicle) to reach a given destination. It is to be noted that the Park&Ride Server makes use of external sources in order to access complementary data regarding weather conditions and routing options with public transportation.

The operation of specific modules within the CDC platform, especially the Disruption Checker, the Parking Prediction and the Route Optimizer, completes the description of the technical architecture. These modules are linked to the Park&Ride Server via the Interface Layer of the CDC platform (cf. architecture Fig. 1) and their main characteristics are as follows⁴:

⁴ It is to be noted that Fig. 2 is simplified as it does not mention the Interface Layer of the CDC platform.

- the **Disruption Checker** identifies the eventual disruptions that occur on a given route. This identification is made possible by the retrieval and the analysis of available data. In our case, this includes public transportation data regarding incidents (delays, service interruption, and accidents), real-time traffic flow in the considered area and activity on social networks (Twitter posts related to the traffic conditions).
- the **Parking Service** collects records related to the evolution of the occupancy rates concerning parking facilities in a given area. Then by extrapolating this data, it estimates the availability of parking lots so that it is possible to propose the best parking options for the considered area.
- the main functionality of the **Route Optimizer** is to identify the best routes to reach a destination. This operation is performed by evaluating, among the matrix of possible routes to a given destination, the minimum of {XYZ}. X corresponds to the use of a private vehicle (distance, time), Y corresponds to the availability of the associated services (e.g. price and occupancy rate of parking facilities) and as for Z it corresponds to the transit options with public transportation (time, changes, price, walking distances).

According to its original description, the CDC platform operates the internal modules and provides the Park&Ride server with the requested data (historical or real-time) in a normalized standard format.

3.3 Related Business Model

From a business perspective, the use of Park&Ride amounts to considering public transportation as an extension of private means of transport or vice versa. It can

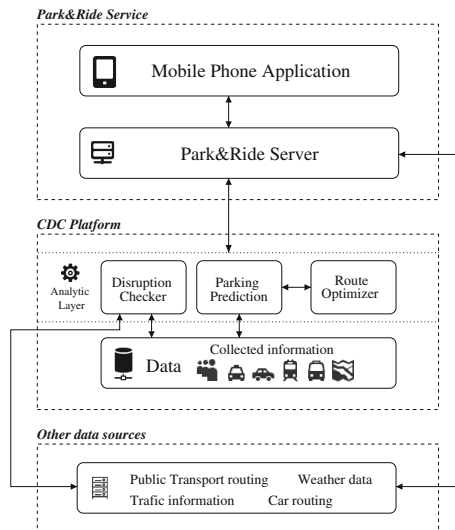


Fig. 2. Architecture of the Park&Ride system

lead to significantly increase the accessibility and the usage of transit networks as well as associated parking areas. Based on this perspective, the most appropriate option is to ensure the development and the deployment of the Park&Ride service through collaborations between public transportation authorities and parking service providers; the main channel to reach the end-user (and “distribute” the service) is the mobile application.

In terms of costs, the resources to mobilize are mainly intended for the deployment and the maintenance of infrastructures for the Park&Ride Server (apart from the development of the mobile application).

In addition, it is to be noted that the Park&Ride service allows suburban/rural users to plan more easily a journey to cities without having to use a private vehicle to go downtown. By leveraging parking information in order to propose public transportation solutions, the system leads to reduce traffic congestion in urban centers. Thus, the demand drops in urban centers regarding the requests for parking. This decrease has the effect of making available more affordable parking options within cities.

3.4 Deployment and Evaluation

The piloting phase of Park&Ride has been carried out in December 2014. It has targeted the city of Helsinki (Finland). The objective was to evaluate the usability and the robustness of the system (supported by the CDC platform) as well as to collect valuable information regarding the feelings of the potential users.

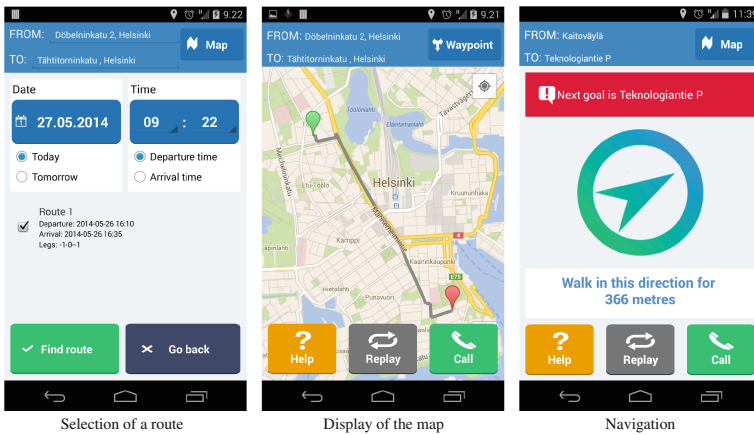


Fig. 3. Screenshots of the Park&Ride application

For testing purposes, a panel of 17 users (an heterogeneous group of voluntary users from diverse backgrounds) was formed among the people working at VTT Technical Research Centre of Finland. These persons were living and working

in the area of Helsinki. Each participant was provided with a car and a mobile device (in this case a Nexus 5 smartphone) running the Park&Ride Android application (its interface is presented Fig. 3). The practical set-up of the tests has led to the following process:

- Each participant had to perform a round trip during about 4 h; the journey included the use of the car, the use of parking facilities and the use of public transportation; the users had the flexibility to choose the destination while the starting point was forced for practical reasons.
- The collection of data regarding the perception of users was done through a questionnaire (details of the aspects to be evaluated are given in the results of the evaluation) and phone interviews.

It is to be noted that the test-related operations on the field have taken place during the week of December 1 (2014).

Figure 4 illustrates the perceptions of users (an average that includes the answers of each participant) concerning aspects such as ease of learning to use the service, ease of use, clearness of the provided information, reliability of the information, usefulness of the service and safety related to the use of the service. On the graph, the blue bars represent the navigation when the user drives the car and the red bars represent the other situations (use of public transportation and walking). The rating of the different aspects of the Park&Ride service was supported by the 5-point Likert scale [11]. The value 1 is associated to the worst case while the value 5 is associated to the best (ideal) case.

The graph shows that, for most aspects, a rating greater than 3.5 (or very close to 3.5) is reached. These results (supplemented by the phone interviews) highlight the fact that the system is perceived as being rather reliable (especially in unfamiliar environments) and they also point out the willingness of the participants to use the service in the future. A notable point concerns the relatively low rating of the clearness for navigation in public transportation and during walking phase (an average of 2.5). An analysis of the responses to the questionnaire and the phone interviews have given an insight for an explanation. It appears

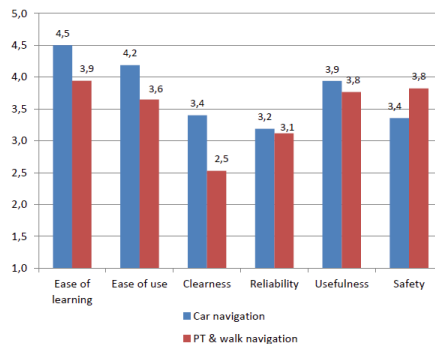


Fig. 4. Evaluation results for the Park&Ride service (Color figure online)

that users, in the context of public transportation, find difficult to understand the flow of information. Due to the various opportunities that are often offered in that phase, the user has to handle a lot of information-rich content.

The evaluation points out the potential of the CDC platform and its suitability for a real urban environment. To further investigate this aspect and complete the previous evaluation, the next section discusses the technical details that have a significant impact on the performance of the CDC platform according to our experience in the CDC project.

4 Return of Experience

In regards to the development and the piloting phase of the platform, we present in this section some reflections and lessons-learned as well as perspectives for the evolution of the CDC framework.

Data Integration. Data integration implied the creation of wrappers for various heterogeneous real-time data sources. The data sources ranged from standardized REST-based APIs to crowd-sourced ones. The development of wrappers is challenging as it requires a precise knowledge of the legacy system for the corresponding data. In our case, the task was especially complex due to the large variation in data types, data models, refresh rates, and storage methods (data dumps vs regular databases). Another challenge was the variation in the quality and the general reliability of crowd-sourced data. For example, during the piloting phase there was several issues in recognizing the right context. E.g. word “Metro” in a Twitter message could mean underground train or downtown as a generic area. Further exploration of this issue is planned.

Presentation and Storage of Data. Designing the object model for the platform was very challenging due the many and varied data sources. It was necessary to normalize certain common attributes, such as the timestamp, which seemed to be different in every data source. We have decided to utilize the ISO 8601 as the baseline for timestamps. We have then implemented converters for each encountered format so that we have the capability to compare, sort, and filter data based on time and date.

Data Provision. During the tests with the Park&Ride mobile application, the most encountered source of problems was the location inaccuracy. The combination of location uncertainties and variations of time lags (because of delays originating from low-quality data links) decreased the quality of the service provided to the users of Park&Ride. Indeed, the handling of multimodal junctions was tricky: How to detect when the user is at the stop or when he has boarded a vehicle? Is the user in the right vehicle? When is the user supposed to get off? In all these cases the location accuracy must be less than ten meters and the real-time data delay (e.g. buses locations) needs to be less than 10s. These requirements were quite often missed especially in the city center with its “urban canyon” environment. We have thus put a lot of effort on the analysis of the reliability and the accuracy of data sources. We have then been able to provide an

estimate that can be injected as metadata, hence enabling the notification of end-users regarding the reliability of information.

Based on the feed-back from developers of mobile applications, we have noticed that the standardized access to real-time data from various sources provided by a centralized platform (i.e. the CDC framework) appears to be valuable. However, there is a trade-off regarding the data when considering the level of processing to perform within the platform. Indeed, raw real-time data has low delays but it requires a lot of data transfer and storage. As for processed data, it is in more compact form but it may limit the potential applicability for developers. The amount of real-time data collected every second by the CDC platform was huge, the storage and the direct distribution of all this raw data was therefore out of question.

Ecosystem of Data Provision. This aspect concerns the adequacy of the CDC platform with the ecosystem related to mobility information. As such, the main issue lies in the fact that data providers are generally local with specificities according to the considered city/area and their systems are (very) proprietary. The data wrapper layer was made as adaptive as possible so that it is relatively easy to implement the integration of new data wrappers to the platform. It was also designed to be lightweight. This approach should simplify the extension of the platform to new cities/areas.

Unreliability of Data Providers. As most of data providers were “loosely-coupled” to the platform, there was no service level agreements (i.e. the supply of data was free of charge). As a consequence, many (unannounced) downtimes occurred. Obviously, downtimes may have cascading effects on other services and their applications. In addition the changes in protocols (e.g. formats of raw-data and interfaces to access data) or even regarding licenses for data usage may happen with short notice. This creates a constant burden relatively to the development of data wrappers as they must be monitored and regularly updated.

Next Steps. According to the feedback from third parties, it is necessary to improve the support for application developers. During this phase of the CDC project, the utilization rate was relatively low for the public APIs (despite appropriate documentation and standardization efforts). The aim is then to produce more mature tools (from a technical perspective) as well as enhanced material in order to ease the use of our APIs within mobile applications.

Based on the analysis of data resulting from Park&Ride pilots, we have noticed additional delays (about 2–5 s) in certain situations. These delays are caused by the data propagation through the layers of our platform and it clearly weakened the user experience. We are therefore looking into more “lighter” and distributed approaches so that in delay-critical cases the real-time data could be directly propagated to the devices of end-users (thus by-passing the analytic layer).

Moreover, there is a clear need to integrate information about ticketing and payment options. It will give the possibility to optimize or sort the routing options according to more (relevant) criteria. Also, the integration of other types

of parking places (e.g. satellite parking or road-side parking) in addition to (traditional) large parking facilities would make the system more dynamic in order to better adapt to various contexts.

5 Conclusion

The Connecting Digital Cities project has initiated the development of a data platform intended to provide a valuable aggregate of information related to urban mobility (e.g. traffic, transportation options, crowd-sourced data from social networks). This aggregate is meant to support the implementation of smart mobility services and it is thus made available to service providers through open APIs. The tests that we have carried out in a real urban environment (i.e. city of Helsinki in Finland) have demonstrated the relevance of the platform from technical and business perspectives. These tests essentially involved the deployment of the Park&Ride (multimodal journey planner) service that interacted with the CDC data platform to retrieve consistent mobility data.

However, many efforts are still needed to improve the overall performance of the data platform. As such, the lessons learned in terms of reliability and suitability of the platform have guided the preparation regarding the second phase of the CDC project. This mainly concerns the integration of other types of data (e.g. information about ticketing and payment options as well as details on non-conventional parking places) and the handling of delay-critical situations by directly targeting the devices of end-users when it is relevant. The goal now is to improve the “maturity” of the technical tools in order to increase the use of the public APIs supplied by the CDC platform.

References

1. Data elements and interchange formats - Representation of dates and times. Technical report ISO, International Organization for Standardization (2004)
2. ASSISTANT project, May 2015. <http://www.aal-assistant.eu/>
3. EIRA, May 2015. <https://joinup.ec.europa.eu/asset/eia/description>
4. Smart Urban Spaces website, May 2015. <http://www.smarturbanspaces.org>
5. ULM, May 2015. <http://www.eitictlabs.eu/innovation-entrepreneurship/urban-life-mobility/>
6. Urban population growth, May 2015. http://www.who.int/gho/urban_health/situation_trends/urban_population_growth_text/en/
7. Chaumette, S., Dubernet, D., Ouoba, J., Siira, E., Tuikka, T.: Architecture and comparison of two different user-centric NFC-enabled event ticketing approaches. In: Balandin, S., Koucheryavy, Y., Hu, H. (eds.) NEW2AN 2011 and ruSMART 2011. LNCS, vol. 6869, pp. 165–177. Springer, Heidelberg (2011)
8. Chaumette, S., Dubernet, D., Ouoba, J., Siira, E., Tuikka, T.: Architecture and evaluation of a user-centric NFC-enabled ticketing system for small events. In: Zhang, J.Y., Wilkiewicz, J., Nahapetian, A. (eds.) MobiCASE 2011. LNICST, vol. 95, pp. 137–151. Springer, Heidelberg (2012)

9. Chaumette, S., Dubernet, D., Ouoba, J., Siira, E., Tuikka, T.: Towards an interoperability evaluation process for mobile contactless city service. In: Proceedings of the Eighth International Conference on Digital Society, ICDS 2014, pp. 65–70 (2014)
10. Erl, T.: *Service-Oriented Architecture: Concepts, Technology, and Design*. Prentice Hall PTR, Upper Saddle River (2005)
11. Gwinner, C.: 5-point vs. 6-point Likert Scales. Technical report, Infosurv (2011)
12. Lahti, J., Ahola, J., Ceri, S., Brambilla, M., Celino, I. Zuccal, M.: Empowering digital cities through interconnected services. In: Proceedings of ITS World Congress 2015, October 2015 (to appear)
13. Ouoba, J., Siira, E.: Many faces of mobile contactless ticketing. In: Proceedings of the Second International Conference on Smart Systems, Devices and Technologies, SMART 2013, pp. 93–98 (2013)
14. Siira, E., Heinonen, S.: Enabling the mobility for the elderly: design and implementation of ASSISTANT navigation service. In: Proceedings of 14th International Conference on Mobility and Transport for Elderly and Disabled Persons, July 2015 (to appear)
15. UN-HABITAT and United Nations Human Settlements Programme: *State of the World's Cities 2008/2009: Harmonious Cities*. The state of the world's cities report, Earthscan (2008)
16. Zuccalà, M., Celino, I.: Fostering innovation through coepetition: the E015 digital ecosystem. In: Cimiano, P., Frasincar, F., Houben, G.-J., Schwabe, D. (eds.) ICWE 2015. LNCS, vol. 9114, pp. 625–628. Springer, Heidelberg (2015)