2015 IEEE Symposium on Service-Oriented System Engineering

Personal Mobility Service System in Urban Areas: the IRMA Project

Gianmario MOTTA, Daniele SACCO, Tianyi MA, Linlin YOU, Kaixu LIU
Dept. of Industrial and Information Engineering
University of Pavia
Pavia, Italy
motta05@unipv.it, {daniele.sacco01, tianyi.ma01, linlin.you01, kaixu.liu01}@ateneopv.it

***Abstract*—We present an ongoing research project, namely Integrated Real-time Mobility Assistant (IRMA). IRMA is a software system that targets the personal mobility in a near future scenario, oriented to green, shared and public transports. IRMA aims to be an extensible, easy-to-implement and sustainable modular platform based on the combined use of multiple information sources (crowd, open, social, and sensor data) and on array of value propositions, each serving a class of stakeholders, which include municipality, users, transport providers. Hence, IRMA supports users in the entire lifecycle of mobility, and municipalities and transport providers in the whole cycle of mobility management. IRMA is deployed on both smartphone and web, and is built on a hierarchy of re-usable web services, that are based on SOA/EDA (Service Oriented Architecture / Event Driven Architecture).**

Keywords-smart city; urban mobility; human mobility; mobility integrator

I. INTRODUCTION

The world population is city-based. Most people live in cities: 3.6 billion nowadays and in 2050 68% of the world population [1]. Also, cities consume 80% of worldwide energy production and generate 67% of energy-related greenhouse gases [2]. Finally, 64% of travel kilometers are urban and the travel within urban areas is expected to triple by 2050 [3].

Hence, urban mobility is a challenge. In 2011 the mobility maturity of 66 cities worldwide was assessed by 11 criteria [4], ranging from public transport share to average travel speed and transport-related CO2 emissions. EU had the best regional performance. The estimated cost of traffic congestion is huge, estimated to be about €111 billion per annum for EU member states [5]. So, the mitigation of congestion is a main priority of many cities and governments in the EU.

What should be an ideal scenario? According to Horizon 2020 [6], the European research framework program, transport shall be smart, green, and integrated. However, a smart transport system also requires a smart user, thus a smart mobility (smart user + smart transport). In a smart mobility, mobile phone users should grow from the current 55-60% to 100% [7]. Ageing people and other users (e.g. car drivers), who hardly benefit of smart phone usage, shall take advantage of smart devices (i.e. “things” in the Internet of Things concept) as smart TVs, in-vehicle devices and open source electronic platforms, that enable an integrated mobility information service, anytime and anywhere. The smart mobility market is growing 20% per year, and it is estimated to exceed $100 billion by 2018 [8]. This forecast includes the smart technologies which will be used, and the expenditure for innovation, design consultancy and engineering, infrastructure development and installation, ICT, software and analytics, and automation and control. It embraces all main transport domains, where smart applications provide services as parking management and guidance, real-time travel information, real time traffic management and other applications.

However, we are facing following main challenges in designing such smart mobility systems:

A. *Service integration*

Current smart mobility systems do not offer a unified and integrated approach to support the whole lifecycle of urban mobility. They always use independent approaches for different mobility needs. [11] Furthermore, the integration of cross-organization smart city services always results poor because of the information islands, and costly because of inefficient procedures. [35][36]

B. *Heterogeneous data integration*

Smart mobility systems involve data from various sources, as traffic data from traffic management systems, transport time table and fare data from transport systems, crowd data from citizens, social data from social network, and sensor data from vehicles, parking lots, etc. The key challenge arises from the diversity of such data. Data may be geospatial, environmental, numerical, and, also, may involve a wide range of data holders. Because of the multiplicity of information sources, data models and types, the integration and analysis of these data are difficult.

C. *Citizen and stakeholder involvement*

The engagement of citizens and stakeholders is considered to be a key element for the success of a sustainable urban mobility. [37] Therefore, smart mobility systems shall shift from traffic management to pursuing both the comfort for citizens and the satisfaction for authorities. Citizens shall not only be the requestors of mobility services, but also play a role in provisioning of services. [11] However, cities shall overcome following challenges: (a) information isolation between citizens, city government and other stakeholders, (b) lack of collaboration processes between stakeholders, and (c) lack of a business model that can be beneficial for all stakeholders.

These challenges are the target of IRMA (Integrated Real-time Mobility Assistant) project, which was initiated in 2011 as a system to support travelers, based on proprietary data provided by transport operators [9]. IRMA evolved through various phases, and it was eventually reconceived as an internet-based platform, that supports the numerous and diverse classes of mobility stakeholders through a layered architecture, that uses open source technologies, and is deployed in the cloud. Such version was conceived as a SS (Service System), i.e. a platform that integrates internet services, internet data and internet of things [10]; it has been also proposed for the H2020 call 7.1: “Connectivity and information sharing”.

This paper illustrates such current version of IRMA. The section on scope illustrates the perimeter and the target of IRMA. A review of current applications is given in “State of

978-1-4799-8356-8/15 $31.00 © 2015 IEEE
DOI 10.1109/SOSE.2015.15

Art". A systematic overview of the system and approach is given in the "Overview" section. Key points on data and data sources are illustrated in the core section. Finally, conclusion summarizes the main achievements and sketches the profile of next release.

II. SCOPE OF IRMA

IRMA addresses mobility of individuals, and it aims at a near future scenario, where green, shared and public transports are replacing private, carbon transports. IRMA is a cross-device and cross-platform system that, on one side, enables users to manage a multimodal mobility and, on the other side, supports the authority in analyzing and monitoring efficiency and effectiveness of the mobility [10].

IRMA handles end-to-end itineraries on multiple transports in order to balance time, energy/pollution and cost of mobility. Therefore, IRMA supports users in plotting itineraries and control them en-route. Specifically, itineraries are updated in real-time during the trip, on the actual progression, position and mobility status. IRMA architecture includes applications for mobile and ad hoc devices (e.g. on vehicle devices), and gathers and interprets multiple relevant sources of information, as open data (i.e. timetable, real-time position and delay), navigation data (e.g. TomTom traffic), crowd-sourcing data from user generated content (e.g. feedbacks sent by travelers on mobility), and social data (e.g. Tweets).

IRMA targets the main mobility stakeholders, namely end users, municipality/transport authorities, and transport service providers. End users (including disabled and elderly people) can benefit a real-time assistant and open information on transports. Municipalities and transport providers analyze mobility status in real-time in the urban area and forecast mobility in front of unexpected changes. Additionally, third parties, as tourism agencies, can use open information as a service to perform analyses and offer additional services.

IRMA is a step ahead the current concept of personal mobility systems, that enable new ways to optimize urban mobility while pushing travelers to use public and shared transport. In short, IRMA targets and support sustainable and citizen-centric transport services.

III. STATE OF THE ART

Smart mobility services for individuals need to support multi-modal trip planning and integrate green and shared transports. In addition, all available internet information, namely open data, sensor data, social data and crowd data shall be integrated. By analyzing these data sources, transport providers can not only forward alerts but also simulate their impact. Thus, emerging smart mobility services will support a deeper insight on traffic conditions, context and flow of people and vehicles.

A. Trip planning services

Typical trip planning services receive query including a start and destination locations, and provide multiple options to users using one or more forms of transportation [12]. Google transit provides GTFS-based public transportation trip planning services. However, GTFS only uses static data, e.g. timetables. Google launched Live Transit Update that uses GTFS-Realtime, to get real-time public transport data from transport authorities. ROVER, Moovel, EMBARK and City Navigator also integrate real-time public transit data. However, only 15 transit providers in a few cities (e.g. New York) provide real-time data to Live Transit Update (as of 2013). Because of low availability and coverage, together with high amount of data in updating positions, raw GPS devices is greatly limited in providing real-time data [13]. Szabo et al. proposed a crowd-sensing service architecture based on Extensible Messaging and Presence Protocol (XMPP) to generate live transit feed [14]. Moovit is able to gather such data and analyze pedestrian flow and public transport conditions. However, crowd-sensing faces many challenges such as motivation of passengers in data gathering [16]. In order tackle this problem, local communities for drivers which enable a "collective" driving model were built by Waze. Furthermore, social interaction features and gamification techniques also increase its user participation. [34] However, no public transport data are integrated. Google completed the acquisition of Waze in 2013 and started integrating Waze's data into Google Map, that shall greatly complement its services.

Advanced trip planning services also support multi-modal transport options to travelers and divide the whole route into several sub-routes with different transport modes. Shared transport is a key strategy for reducing fuel consumption and pollutant emissions in the front of the global climate emergency [17]. CityMapper and Roadify integrate real-time bike sharing data (e.g. position of rental stations, available bikes). Moovel and Roadify gather car sharing data (car position, fuel level) from Car2go. However, current trip planning services neglect Park and Ride mode (i.e. travellers drive to transit stations, park their cars there and use public transit to travel) which plays an important role in reducing car use and carbon footprint [17]. Li et al. proposed a multimodal trip planning system with Park and Ride mode and real-time transit data, but this system is in early stage [18].

B. En-route services

En-route re-planning is required with the rise of uncertainty and disruption of transport services. [19] The state of the art services implement three re-routings, respectively based on position, speed, and event. Google transit supports a position based re-routing, but without considering traffic conditions, e.g. traffic jams. Waze proposes a re-planning that considers both events and speed. Indeed, it re-plans routes that skip traffic jams by using the vehicle speed data gathered from users' mobile sensors and the traffic congestion data uploaded from other users on the same route. However, its accuracy reflects the number of users: it is higher with a higher density and lower with a lower density.

C. Infoprmation sources: crowd data and social netowrks

In current applications and researches, data from crowd-sourcing/crowd-sensing and social network are the main trend. As already mentioned, Waze collects both crowd data and sensor data from users' mobile phones. Even if analysis and event detection are on an early stage, researches show a high potential in event detection and forecasting. Fire et al. crawled report data from Waze and processed data for KPIs and accident forecasting. [20] Silva et al. analyzed Waze alerts from Twitter, and found that it could identify traffic problem reasons, that hardly can be spotted by traditional sensors. [21] However, both works are in a concept stage and no practical solution is proposed.

Roadify collects transport status from both crowdsourcing and Twitter. It crawls event data from Twitter by identifying locations and tags for each public transport route. Still, these data are only for static display and are not accurate enough. SMARTY [22] project proposes a social-sensing approach for detecting traffic status, accidents and special events. It uses text-mining technologies to analyze Status Update Messages (SUM). However, the actual result of social-sensing is unknown. Pan, Bei et al. proposed a two-step approach that detects changes of travel patterns and performs text-mining on social networks. In experiment on dataset of Beijing, 86.7% of events are detected. [23]

D. *Integration of mobility services*

According to the previous discussion, current personal mobility applications do not fully integrate services and data sources. Furthermore, they do not provide complete and robust mobility analysis services. Therefore, a platform which gathers, stores and processes information should: (1) Perform analytics to discover and forecast patterns about mobility of people, mobility of vehicles, environment and context; (2) React in real-time to events or changing conditions; (3) Integrate data sources (e.g. using linked data and/or semantic web technologies); (4) Make data easily accessible and reusable by end user services and/or external services. Here below, we review 5 integrated smart mobility projects with a higher service and/or data integration.

MOBINCITY project provides Fully Electric Vehicles (FEV) related services. The key innovations is the integration of mobility services, FEV charging scheduling services, and a complete service platform for FEV drivers. However, this project focuses on services for FEV users, and only sensor data are considered.

OPTICITIES project aims at providing multimodal solutions, that deliver stakeholder-oriented services and supports individual mobility such as en-route trip planning. It proposes a standardized data exchanging solution for multi-modal urban mobility and enables mobility analysis, prediction and decision making for traffic control. However, crowd data and social data are not addressed.

Instant Mobility, a FP7 funded project, integrates data from transportation systems and sensors. Individual mobility services that support real-time trip planning are implemented. Social data are used for analyzing and predicting traffic conditions by transport authorities. However, crowd data are only partially used for ride-sharing.

MyCity platform provides context-aware mobility services into several mobile/web/on-board applications in a pay-per-use transactional model. It contains a cloud-based Context Broker Platform (CBP) which gathers and analyzes real-time travel data for event alerts and action suggestions. However, it does not integrate social data into the platform.

Siemens mobility solution covers real-time multi-modal trip planning, parking, reservation and ticketing services. A business-to-business platform for service providers is built to extend their services in an easy, centralized and standardized way. Again, like other integrated solutions, Siemens mobility does not consider the use of data from social network.

In short, most of these integrated mobility projects support multi-modal mobility services and provide a wide range of integrated transport information. However, these projects do not fully integrate available mobility data (i.e. crowd data and social data) and the provided services are isolated, thus being not easily reused or integrated. Finally, their service coverage is wide and complex. This may be a barrier to dissemination.

IV. OVERVIEW

A. *Scope of an ideal mobility integrator*

Public and collaborative transport are essential to an urban sustainable mobility. An integrated transport can lead to an improved quality of life, socio-economic development and urban renewal. This integration also enables higher density development with positive benefits on the environment: more efficient land use, higher energy efficiency, reduced pollution, climate change mitigation and protection of open space through smarter growth pattern [24] [25].

A future, ideal scenario requires a new mobility service concept. Mobility Integrator is an innovative approach that integrates Information & Planning, Transport Services, Infrastructure and Traffic Management. In short, it shall provide a smooth and convenient integrated mobility platform. In a technological perspective, a mobility integrator is the integration of various systems, namely Traffic Management Systems, Transport Systems, User Systems, and Vehicle Systems.

Within the Mobility Integrator concept let us address the personal mobility, that is the focus of our IRMA project. An ideal future system should support all stages of personal mobility (Fig. 1), as (1) planning (i.e. plotting the trip end-to-end), (2) execution (i.e. en-route phase and related transactions), (3) monitoring (i.e. analysis of information on current and forecasted conditions that affect the selected itinerary), (4) control (i.e. identification and selection of an alternative route in front of condition changes).

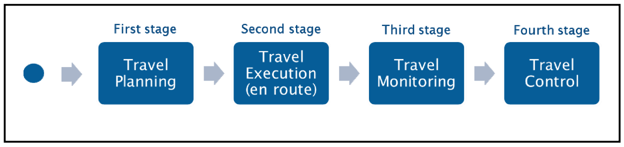

Figure 1. Mobility stages.

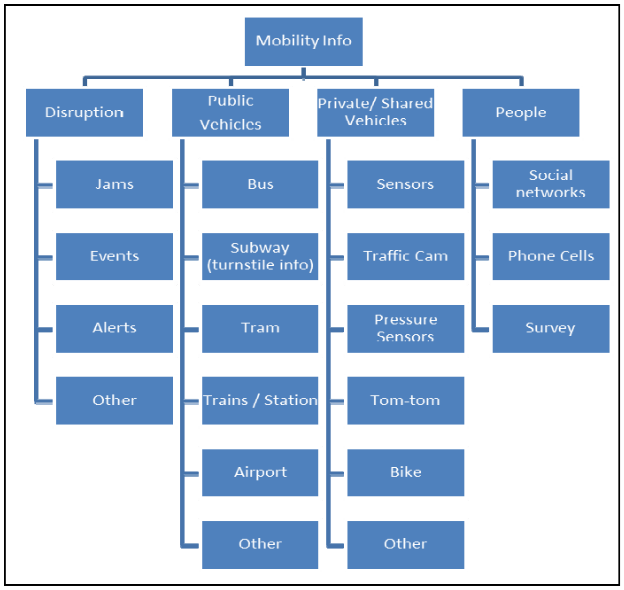

Figure 2. Potential information sources in mobility.

B. Information integration

A key point for mobility integration is information completeness. Information sources are multiple and diverse. On one side, in Internet of Things (IoT) [26], multiple layers of information are being created: advanced mobility systems introduce V2V (Vehicle to Vehicle), V2I (Vehicle to Infrastructure), I2I (Infrastructure to Infrastructure) layers [27]. On the other side, social networks [28] and open data [29] are multiplying potential information sources on Internet.

Such multiplicity of information source, that we informally show in Fig. 2, implies that a complete and sustainable system should be able to (a) access all relevant information sources through web services (b) integrate information of different sources. To integrate mobility information we adopted a simple layered framework. The layered comprises both integrated information and information sources. Integrated information is structured according to integrated schemas.

Figure 3. Potential information sources in mobility.

IRMA uses a multi-layered mobility service design framework that is described in [10] (Fig. 3). At the lowest layer, namely "Service 1: Urban Trip Planner", a generic city map (e.g. Open Street Map) is integrated with the information about transports, as trains, trams, subway, buses, etc. At the upper level "Service 2: Alert Management" information of level 1 is integrated with dynamic information about events as traffic jams, shows, etc. A further set of service create the applications i.e. services to users. E.g. Alert Management includes services that display, analyze, and simulate alerts. Services to users conceived to support all stakeholders of individual mobility, as travelers, authorities, and transport providers. The services can be accessed by web, smart phone, smart TV and other special devices.

Thanks to this separation of information integration from services to user, the same information source/aggregation can be used by multiple services. The related design framework is described in a previous paper [10].

C. Services to user: Mobiity Analyzer

In order to exemplify IRMA's services, we here consider information supporting mobility analysis, forecasting and assistance.

Mobility analyzer stores a mobility map and related mobility data. The mobility map describes mobility resources within the urban area, by route, time, and mode. In turn, mobility data describe the mobility load by route, time and date. Users can analyze mobility across any urban area, while municipalities and transport providers can analyze actual versus expected mobility performances. These historical data are an input to the mobility forecasting service.

Fig. 4 shows the mobility analysis layers: the first layer represents a static map of an urban area showing streets and transport network. The second layer highlights the routes and stops of the transports. The third layer shows the scheduled position of vehicles according to their timetables. The user, through this layer, can have a preliminary assessment of the ideal performance of the mobility. The fourth layer indicates the real-time position of vehicles. The last layer gives the Key Performance Indicators (KPIs) on mobility.

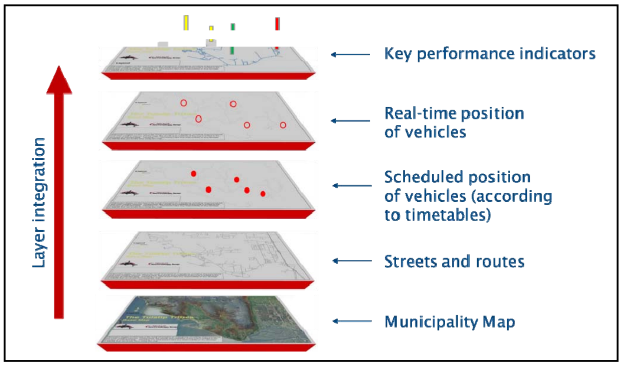

Figure 4. Mobility analysis layers.

Performance measurement address the gaps between current and desired performance and shows the progress in closing gaps. KPIs identify precisely where to take action to improve performance. Mobility performance analysis includes: (a) Cost indicators, that assess the unit costs of input and output and the productivity and utilization rate of the resources, e.g. average travel time and average cost to customer; (b) Quality indicators, that measure the consistency of process input and output in term of compliance, customer satisfaction and availability, e.g. stop overcrowding, urban area overcrowding, quality perceived and network condition; (c) Service indicators, that measure the time aspects of the service in term of service duration, punctuality, flexibility and fulfillment, e.g. delay at stops, path reliability, delay forecasting efficiency, land and social inclusion, average speed of vehicles.

D. Services to user : Mobility Forecaster

Forecasting mobility of objects (e.g. people) compromises mining and machine learning of a sequence of location and timestamp elements. So, this kind of data have been increasing rapidly with the development of GPS, RFID, sensor, wireless, and video technologies and few tools are available for mining mobility. Mobility forecasting needs the integration of many functions including pattern mining and trajectory mining based on state of the art methods. The tool that can be provided to data analysts shall automatically detect approximate periods in movements, collective movement patterns, and perform trajectory clustering, classification and outlier detection for geometric analysis of trajectories. In such perspective, data analysts will be able to answer fundamental questions, e.g. what are the frequent patterns of people's travels? How big attractors and extraordinary events influence mobility? How to predict areas of dense traffic in the near future?

Data analysts will be able to work on large data sets by exploiting Big Data technologies. Stream processing is another fundamental task within mobility forecasting. Models must consider real-time events and exceptions that may change forecasting results (e.g. an unexpected traffic jam). These events must be managed in real-time and complex event processing correlate them with analytics outcomes. Forecasting results correlated with real-time events work as input to the mobility assistant, where the user can choose the best route, taking into account the dynamics of transportation modes and user constrains.

E. Services to user :Mobility Assistant

It shall assist the end user to plan, configure, monitor, and reschedule mobility across multiple modes. It manages and supports the mobility itinerary by two phases in the mobility life cycle with different services.

Before the trip, a trip planner processes user's mobility request that may concern an individual trip or a calendar. The trip planner defines the optimal mobility plan by accessing timetables, real-time delay, events and mobility forecast. The user will choose the ideal itinerary. Transport time is based on the forecaster that adjusts the standard time on the traffic load projections related to transport modes, the specific daytime, date, and the route from A to B.

During the trip the user receives information about disruptions (Event Notifier) and choose a viable alternative (Compensation Engine). This scenario is shown in Fig. 5. Event Notifier is based on message pushing services and it is a set of instances that are activated when the user confirms and actually starts the trip. It concerns all connections of the itinerary and processes relevant event information provided by the mobility analyzer. A Compensation Engine processes mobility alternatives in front of a disruption or a change, by browsing alternative options (as an alternative the request handler could fetch a plan B in advance).

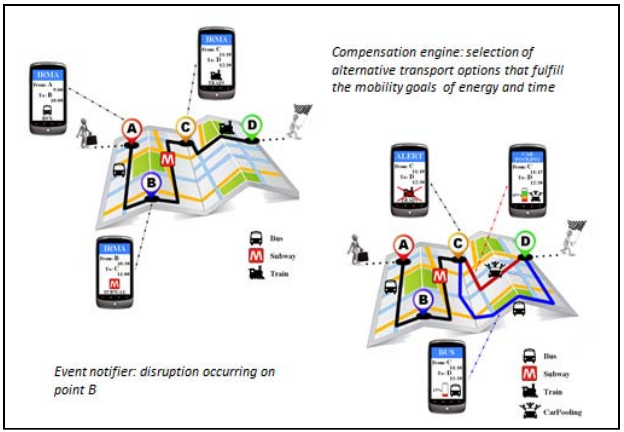

Figure 5. During the trip - Reschedule a trip - illustrative screens.

Furthermore, citizens becomes also an information producer because he/she can identify potential issues relative to mobility or network condition (e.g. potholes, unexpected strikes, delays not reported), spot them with a picture that is automatically geo-tagged in a map. Geo-tagged tweets about these issues are also event data sources. Citizens and municipalities can also help to collect accessibility data all over the city. These data will be used by travel guide services for wheelchair users.

F. Relations among stakeholders

To provide our services we can exploit a multisided tech-enabled business model, as depicted in Fig. 6.

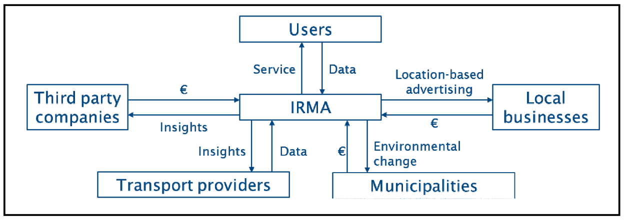

Figure 6. Business model.

IRMA services are provided on the cloud. By providing IRMA as a service to the end user, we can benefit from the flexibility to let demand drive consumption so that resources can be rapidly and conveniently provisioned. Also our IT infrastructure is totally on the cloud. Thus, we can leverage the same benefits (e.g. service level management, auto-scaling) and focus only on operating expenditure.

According to McKinsey, multisided business models create value through interactions among multiple players rather than traditional one-on-one transactions or information exchanges. [31] In a multi-sided business model each interaction creates value to both parties where one party represents always a "collector" (i.e. IRMA, in our case). On the user perspective, we offer a service on a free (or freemium) basis in change of data collection about user's (anonymized) position and crowd-sourced feedbacks; Transport providers are pushed to publish open data in order to receive insights and analytics about the transport service they are providing; Municipalities can benefit of environmental change and urban area sensing and pay for such insights. Offering also data as a service, by providing API to analyze geo-located aggregated data about mobility of people. This API can be accessed on a pay-as-you-go basis, measured by number of queries; Local businesses can benefit of IRMA by implementing location-based marketing campaigns that address specific users during their mobility in the urban area.

V. KEY ASPECTS

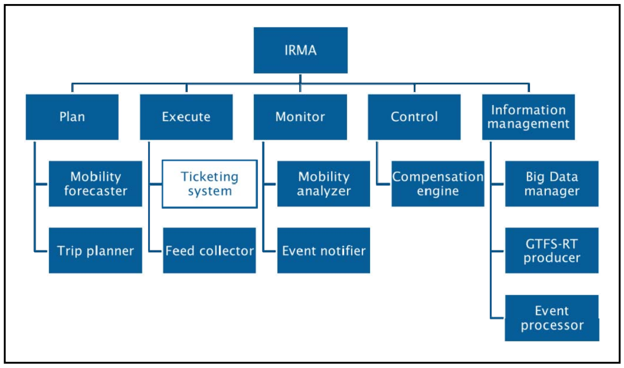

Figure 7. Functional decomposition

In order to provide a proof of concept we illustrate the modules composing the architecture of our system. A functional decomposition diagram, as shown in Fig. 7. To access data a

SOA-EDA (Service Oriented Architecture–Event Driven Architecture) has been adopted. Mobility services that provide assistance, analysis and forecasting need support by ad hoc subsystems.

Fig. 8 shows the overall architecture of IRMA. Next sections give details about the data sources and implementation technologies used in data collection, storage and processing.

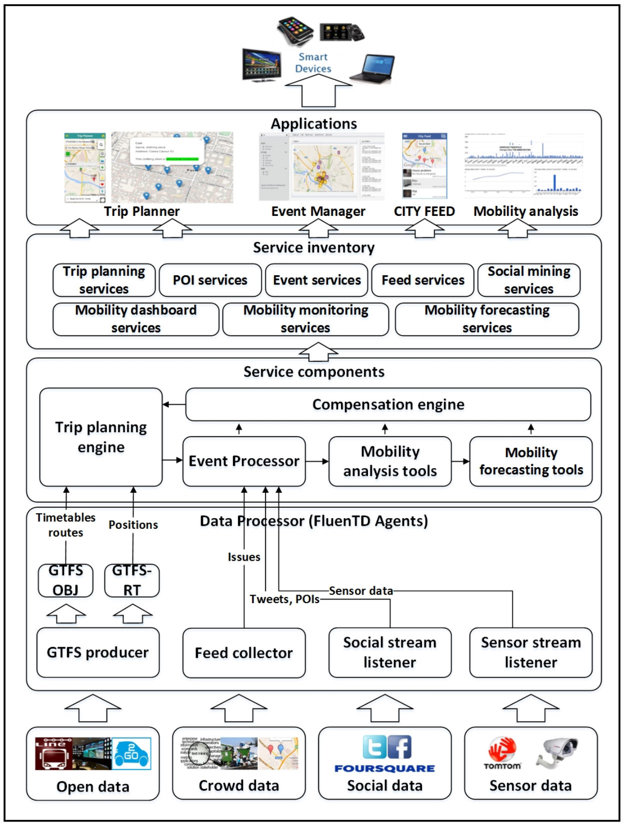

Figure 8. Overall architecture.

A. *Data availability*

Access involves a wide range of data sources: (1) Open data: timetables and static data of trains, buses, underground and alike; real-time vehicle data from buses, trains, etc.; availability of sensors to track shared transport means (shared bikes, cars, taxi) and parking lots; traffic data from traffic management systems; (2) Crowd data from users and/or their devices. (3) Social data from location based social networks.

The key challenge arises from the variety of such data. Data may be geospatial, environmental, numerical, and, also, involve a wide range of data holders. Because of their diversity, a universal integrated access may result hard: transit data are often locally stored, thus creating a patchwork of repositories; legacy systems create lock-ins, that prevent adoption of open standards and hamper interoperability; many incumbent service providers, particularly when relying on sales of data, regard exclusive access to transport data as a competitive advantage; the differences among countries, regions and transport modalities in terms of data format, development stage, market maturity and business models, prevent universal solutions; data quality, security and privacy issues in heterogeneous data.

1) Transit data

Transit data can be easily integrated by open data initiatives. Open data initiatives require that public sector data are freely available to everyone. Public transport data are open data. According to European Union, transport open data include geospatial data (e.g. bus stops, routes), timetables, fare cost, road closure, road network with metadata (speed limitation, addresses, etc.), real-time or archived traffic data, accident data, historical reports and cultural heritage (transport organizations exist for a long time and have big cultural heritage databases including pictures).

General Transit Feed Specification (GTFS) is an open and common data format for temporal schedule and spatial data that is used for transit trip itinerary planning. Over 400 transportation agencies are publishing their schedules and other information using GTFS format. GTFS is a standard de facto because of its wide spread and, thus, a tool designed for one city/agency can be applied elsewhere. GTFS has also been specified for real-time information and currently supports trip updates (i.e. delays, cancellations, changed routes), service alerts (stop moved, unforeseen events affecting a station, route or the entire network), vehicle positions and congestion level. Thus, users can benefit of real-time information about disruptions, location of vehicles, delays and expected arrival times.

2) Sensor data

As stated in [32], sensor networks can be defined as large-scale ad hoc networks of homogeneous or heterogeneous, compact, mobile or immobile sensor nodes that are randomly deployed in an area of interest. [33] observes that recent advances in miniaturization and low-cost, low-power design have led to active research in large-scale, highly distributed systems of small, wireless, low power, unattended sensors and actuators. Ad-hoc deployable, wireless sensor networks can observe the environment in a fundamentally different way than previous systems close to the phenomena in question, over a wide area, and densely in both time and space. They may succeed in applications where traditional solutions have failed.

Different types of data are collected by the sensor nodes. This includes application specific environmental parameters as well as generic data such as meteorological or differential GPS. These data can be in different forms, digital and analogue, spatial and temporal, alphanumeric or image, fixed or moving. Here we consider sensor data coming from external sources because we could not deploy our own sensors in the urban area. Specifically we consider data in terms of real-time traffic data, geocoding engine, traffic camera parking and shared transport (e.g. Car2go services).

3) Crowd data and social data

To develop a vision of an explicitly sustainable transport system which has the power to motivate people to change, it is necessary to combine together substantial technology and behavioral changes, and to ensure ecology, economic viability and good quality of life. Technology change ensures that the remaining transport will be sustainable. The optimization and intelligent design of multimodal mobility concepts via information and communication technologies also play an important role [32].

Last years the role of the user on the web has shifted from consumer to producer of information. Web 2.0 changed users' approach to information creation and exploitation. The paradigm also shifted to social. Web has become an essential need for people, a sort of commodity that is used to communicate, interact, share information and even maintain relationships thanks to social networks.

The last advent of smartphones, equipped with GPS sensors allowing users to geo-locate themselves, can take to the next shift, from a social and collaborative Web 2.0 to a local and mobile Web 3.0. One among the first achievements has been the integration of Geographic Information Systems (GISs) and social networks resulting in new location-based capabilities. Social networks that include location information into their contents are called Location Based Social Networks (LBSNs). Thus, LBSNs offer spatio-temporal information which can be accessed through public Application Programming Interfaces (APIs) and draw the interest of researchers with diverse scientific backgrounds. This availability of data enables a potential use of geo-located content as an additional, low cost and infrastructure-less source of information for urban sensing in Smart Cities.

Thus, we propose 2 approaches: (1) Crowd-sourcing. Citizens identify (or confirm) issues during their mobility by providing a brief description and a picture taken with their mobile. Feeds are tagged according to a predetermined list of issues (e.g. delay not reported, unexpected strike and infrastructure). When the feed is uploaded to the system also its geo-location is stored. (2) Social-mining: to define activity areas in a city we can model crowds by density clusters of geo-tagged tweets, and then drill down them according to different scales. We can also model crowds next to transit stops. The social-mining activities also contain event detection that are using models trained from crowd-sourcing data.

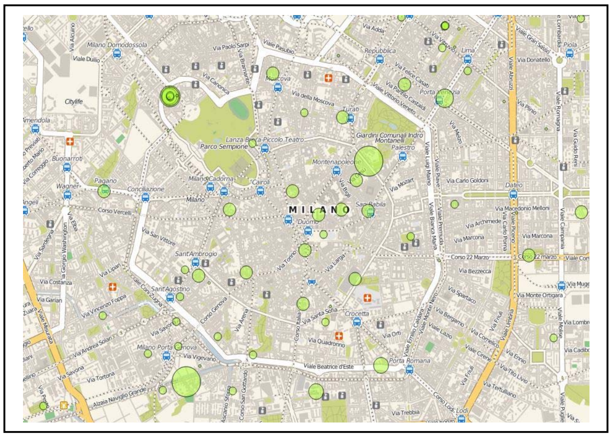

Figure 9. Social data clustering.

B. Data collection, storage and processing

Data processor (FluenTD agents) is implemented by Node.js on cloud for real-time data collection, filtering and processing. Viable alternative is Python and we used it in conjunction with SAP Hana to provide a validation of social data mining discussed in previous sections, as shown in Fig. 9. Since Node.js is very light-weight and scalable, we are able to retrieve and process social data in real-time for mobility event detection. Following data are processed: (1) GTFS data stored in Amazon S3 as OBJ files that describe graphs (i.e. routes) on OpenStreetMap; (2) GTFS-RT data coming in near real-time (i.e. typically an update every 30 seconds) from the GTFS-RT producer that transforms data from specific data format (e.g. XML, JSON) to GTFS-RT; (3) Sensor data about traffic coming from TomTom API in near real-time (i.e. an update every 60 seconds); (4) Real-time tweets by exploiting Twitter Stream API; (5) Crowd data coming from clients of CITY FEED system.

Big Data Manager is a set of technologies exploiting Amazon Web Services and Qubole features. GTFS-RT Producer, Feed Collector, Stream listener for social media data and Tomtom API, and Event Processor (implemented by Node.js) are implemented and is deployed on OpenShift cloud services. Data are preprocessed and stored for further analysis. In order to analyze data, we use Qubole cloud service to define models and set parameters, that is a Hadoop-as-a-service solution that implements most out of the analytics capabilities offered by Hadoop platforms (e.g. Pig, Hive). Qubole has been designed to work on Amazon S3, so integration between the services is built-in. Thus, storage and analytics happen on the cloud without the need of high capability expenditures. Mobility analysis components retrieve analytics results and can be correlated with real-time events retrieved by the Event Processor in Node.js. Output is sent directly to the mobility services. Services (e.g. trip planning services, event services, POI services, feed services, mobility analysis services.) are registered in a service inventory.

C. Data distribution

Data distribution happens during the provision of the mobility services. All previous phases involve data preparation in order to meet stakeholder needs within data distribution phase.

Mobility assistance is provided by implementing a back-end application and a series of applications for different platforms (e.g. web, smart phone, smart TV and on-board devices). Mobile platform are fundamentals because the user needs to track, monitor and receive notifications about his/her trip. The back-end is based on OpenTripPlanner (OTP) that supports GTFS and multi-modal trip planning. Following applications were developed: (1) En-route event awareness trip planning application based on A* algorithm; (2) CITY FEED: a crowd-sourcing based collaborative application for mobility/infrastructure issue reporting and fixing. [30] CITY FEED is also used for collecting accessibility data for disabled and elderly trip planner. These accessibility data will help trip planner to provide accessible route for wheelchair users; (3) Event Manager used for publishing and monitoring events.

Mobility analysis and forecasting services are offered by using a different back-end developed in Node.js. It is necessary in order to manage vast amounts of data shown on maps developed by using Leaflet framework. Mobility analysis and forecasting services will be accessible by Web Services in order to provide smarter services for ad hoc devices, e.g. applications for tablets and smart TVs to monitor KPIs in bars close to important stops. Mobility analysis and forecasting services contains (1) mobility dashboard service that covers all mobility indicators and provides a territorial intelligence that integrates GIS and BI systems. (2) Mobility monitor that shows delays and other basic indicators on specific transit stops. It can be installed in coffee shops, bars, building lobbies, stops themselves, by exploiting tablet, smart TV or ad hoc devices built with open source electronic. Mobility monitor offer an on-the-spot service to users who do not own mobile devices and elderly people. (3) Mobility forecaster that involves the development of a Big Data analytics platform that can be accessed by a web application to set parameters for the prediction models. The platform serves data analysts from municipalities and/or transport providers.

VI. CONCLUSION

First prototypes are available as follows:

- Android App for mobility assistance in Pavia (Italy): http://camellia.unipv.it/servizi/tmp/tripplanner.apk
- Web application for mobility assistance in Pavia (Italy): http://tripplanner-irmapavia.rhcloud.com/otp-leaflet-client/
- Android App for CITY FEED in Pavia: http://camellia.unipv.it/servizi/tmp/cityfeedapp.apk
- Web application for CITY FEED in Pavia: http://cityfeedpavia-irma.rhcloud.com/
- Web application for mobility analysis (mobility simulator and land inclusion) in Pavia (Italy): http://mobilitymap.unipv.it
- Web application for mobility analysis (real-time mobility status) in Boston (USA): http://mobiboston-robolab.rhcloud.com/

IRMA is expected to impact environment, business and users. As far as environment is concerned, the service will alleviate congestion and reduce pollution levels, since it will support travelers in using public/shared transport systems even with complex connections and real-time information. Also, a variety of stakeholders can benefit from IRMA. In the business perspective, municipalities and transport providers can analyze and forecast mobility in terms of time, route, connection, and mode, and, therefore, can optimize transportation resources. Thus, a more efficient use of existing transport infrastructure becomes manageable. Finally, mobile users can define their trips and actually travel in an optimal way, without any previous knowledge of the transport system.

Table I shows a heat-map of the most important existing services/applications/systems in terms of data sources, services and terminals. We consider them regardless the final goal of the service (e.g. public transport, car-sharing, navigator, etc.) in order to understand which option may be able to scale up and provide a complete service to travelers, i.e. our scope, as shown in the final row.

IRMA is feasible because it mainly integrates existing technologies and reuses open source software; moreover, it does not require substantial investments, since its cost for deployment is an operating expenditure (i.e. PaaS cloud solutions). On the other side, IRMA requires designers and implementers who are able to integrate disparate systems.

Generally, IRMA is beyond the state of the art, especially in data integration. Specifically, it defines a framework for ITS deployment based on consolidated knowledge about Big Data and Complex Event Processing, and thus accelerated roll-out of the related services and technologies to integrate heterogeneous sources of data. Moreover, need of deployment of our system may foster open data initiatives in smart cities, by unlocking the potential of vast amounts of transport data. Collection and processing of related big data can result in insights that can be potentially sold to local businesses in order to run location-based advertisement and marketing initiatives. IRMA will enable transport providers to exploit maximum capacity from their existing systems and networks, facilitate the obtaining of the quality and type of data needed by decision makers to make performance-based investment decisions and cost benefit analyses.

Given the variety of data in IRMA, many challenges in data integration are to be faced, on data reliability, accuracy, duplication, security and privacy preserving. Solving these issues are our priority missions in next stage:

A. Data integration and data quality

1) Data de-duplication

Social data and crowd data contain not only text, but also images and geo-locations. Therefore, duplicated data hardly can be identified not from the bare text, but also consider image and geo-location data. Combinations of technologies as text mining, image similarity and record linkage will be used for such data de-duplication.

2) Data reliability assessment and validation

To evaluate reliability of events detected/reported from social network and crowd, comprehensive evaluation methods will be used, .e.g. PageRank based influence evaluation on authors.

B. Data security and privacy preserving

Services of IRMA are deployed on hybrid cloud which stores citizen's data locally while provides public mobility services on public cloud. Data processing between local and public cloud will be strictly controlled. Object matching techniques will also be used for privacy preserving in image data. E.g. human face on images (from crowd or social data) will be detected and masked automatically.

REFERENCES

[1] G. K. Heilig, "World urbanization prospects the 2011 revision," United Nations, New York, NY, 2012.

[2] D. Hoornweg, M. Freire, M. J. Lee, P. Bhada-Tata and B. Yuen, "Cities and climate change: responding to an urgent agenda," 1st ed., Washington, DC: World Bank Publications, 2011.

[3] T. Dixon, "Sustainable Urban Development to 2050: Complex Transitions in the Built Environment of Cities," WP2011/5, 2011.

[4] W. Lerner and F. J. Van Audenhove, The future of urban mobility: Towards networked, multimodal cities in 2050, Public Transport International-English Edition, 61(2), 14, 2012.

[5] P. Christidis, J. N. Ibáñez Rivas, "Measuring road congestion," Institute for Prospective and Technological Studies, Joint Research Centre, Brussels, 2012.

[6] Horizon 2020: http://ec.europa.eu/research/horizon2020/

[7] Global mobile statistics 2014 Part A: Mobile subscribers; handset market share; mobile operators: http://mobithinking.com/mobile-marketing-tools/latest-mobile-stats/a#smartphone-shipments

[8] Smart Transportation Market worth $104.19 Billion by 2019 : http://www.marketsandmarkets.com/PressReleases/smart-transportation.asp

[9] G. Motta, D. Sacco, A. Belloni and L. You, "A system for green personal integrated mobility: A research in progress," in 2013 IEEE Int. Service Operations and Logistics, and Informatics Conf., pp: 1 - 6.

[10] Motta, Gianmario, et al. "Mobility Service Systems: Guidelines for a possible paradigm and a case study." Service Operations and Logistics, and Informatics (SOLI), 2014 IEEE International Conference on. IEEE, 2014.

[11] A. Sassi, F. Zambonelli, "Towards an Agent Coordination Framework for Smart Mobility Services," presented at ACM Int. Workshop on Agents in Traffic and Transportation, Paris, France, 2014.

[12] H Bast, et al., "Transit routing system for public transportation trip planning," U.S. Patent 8,417,409. Apr. 9, 2013.

[13] P. Brosi, "Real-Time Movement Visualization of Public Transit Data," M.S. thesis, Faculty of Engineering Department of Computer Science, Univ. of Freiburg, Freiburg, Germany, 2014.

[14] R. Szabo, E. Karoly, and W. Bernát. (2013). Measurements of a real-time transit feed service architecture for mobile participatory sensing. Presented at Wireless Days (WD), IFIP. IEEE.

[15] K. Farkas, A. Z. Nagy, T. Tomas, R. Szabo, "Participatory sensing based real-time public transport information service," in IEEE Int. Conf. on Pervasive Computing and Communications Workshops (PERCOM Workshops), 2014, pp. 141-144.

[16] R. K. Ganti, Y. Fan and L. Hui. (2011). Mobile crowdsensing: current state and future challenges. IEEE Commun. Magazine, 49 (11), pp. 32-39.

[17] Z. Liu and M. Qiang. (2014). Bus-based park-and-ride system: a stochastic model on multimodal network with congestion pricing schemes. Int. Journal of Systems Science. 45 (5), pp: 994-1006.

[18] J. Q. Li, et al., "A multimodal trip planning system incorporating the park-and-ride mode and real-time traffic/transit information," in Proc. ITS World Congress, 2010, Vol. 25, pp. 65-76.

[19] A. Kramers. (2014). Designing next generation multimodal traveler information systems to support sustainability-oriented decisions. Environmental Modelling & Software 56, pp. 83-93.

[20] M. Fire, D. Kagan, R. Puzis, L. Rokach and Y. Elovici, "Data mining opportunities in geosocial networks for improving road safety," presented at the 27th Convention of. IEEE Electrical & Electronics Engineers in Israel (IEEEI), 2012.

[21] T. H. Silva, et al. (2013). Traffic Condition Is More Than Colored Lines on a Map: Characterization of Waze Alerts. Social Informatics. Springer International Publishing. pp. 309-318.

[22] G. Anastasi, et al. (October, 2013). Urban and social sensing for sustainable mobility in smart cities. Sustainable Internet and ICT for Sustainability (SustainIT), pp. 1-4.

[23] B. Pan, et al., "Crowd sensing of traffic anomalies based on human mobility and social media," Proc. of the 21st ACM SIGSPATIAL Int. Conf. on Advances in Geographic Information Systems, 2013, pp. 334-343.

[24] P. Ribeiro, J. F. G. Mendes. (2010). The role of public transport in the achievement of urban sustainable mobility in mid-sized cities. Presented at Congresso Luso-Brasileiro para Planejamento Urbano, Portugal.

[25] D. A. Hensher, G. Wong, “Different Approaches to Public Transport Provision,” LTA Academy-Journeys, vol.31, 2011.

[26] L. Atzori, A. Iera and G. Morabito. (2010). The internet of things: A survey. Computer Networks. 54 (15), pp: 2787-2805.

[27] V. Milanés et al. (2012). An intelligent V2I-based traffic management system. Intelligent Transportation Systems, IEEE Trans. 13 (1), pp: 49-58.

[28] H. Kwak et al. “What is Twitter, a social network or a news media,” ACM Proc. of the 19th Int. Conf. on World wide web, 2010, pp. 591-600.

[29] C. Hartung et al., “Open data kit: Tools to build information services for developing regions,” in Proc. of the 4th ACM/IEEE Int. Conf. on Information and Communication Technologies and Development, 2010, pp. 18.

[30] Motta, Gianmario, et al. "CITY FEED: A Crowdsourcing System for City Governance." Service Oriented System Engineering (SOSE), 2014 IEEE 8th International Symposium on . IEEE, 2014.

[31] G. Motta, A. Ferrara, D. Sacco, L. You and G. Cugola. (2013). Integrated Mobility: a research in progress. Journal of Software Engineering and Applications. 6 (3), pp. 97.

[32] I. Gallegos, A. Q. Gates and C. Tweedie, “DaProS: A Data Property Specification Tool to Capture Scientific Sensor Data Properties,”, ER'10 Proc. of the 2010 Int. Conf. on Advances in conceptual modeling: applications and challenges, 2010, pp. 232-241.

[33] J. E. Elson, “Time Synchronization in Wireless Sensor Networks,” Ph.D dissertation, University of California, Los Angeles, US, 2004.

[34] You, Linlin, et al. "Social data analysis framework in cloud and Mobility Analyzer for Smarter Cities." Service Operations and Logistics, and Informatics (SOLI), 2014 IEEE International Conference on. IEEE, 2014.

[35] David, Bertrand, Chuantao Yin, Yun Zhou, Tao Xu, Bingxue Zhang, Huiliang Jin, and Rene Chalon. "SMART-CITY: Problematics, techniques and case studies." In Computing Technology and Information Management (ICCM), 2012 8th International Conference on, vol. 1, pp. 168-174. IEEE, 2012.

[36] Pan, Jin-Gu, Yu-Fan Lin, Su-Yi Chuang, and Yu-Chia Kao. "From governance to service-smart city evaluations in Taiwan." In Service Sciences (IJCSS), 2011 International Joint Conference on, pp. 334-337. IEEE, 2011.

[37] Lindenau, Miriam, and Susanne Böhler-Baedeker. "Citizen and Stakeholder Involvement: A Precondition for Sustainable Urban Mobility." Transportation Research Procedia 4 (2014): 347-360.

TABLE I. HEAT-MAP OF EXISTING SERVICES/APPLICATIONS/SYSTEMS

	Data collection				**Data processing**			**Data availability**		
	Transit data	***Crowd data***	***Sensor data***	***Social data***	***Assistance***	***Analysis***	***Forecast***	***Fixed***	***Mobile***	***Ad hoc***
Sidecar	🟥	🟩	🟥	🟥	🟩	🟥	🟥	🟥	🟩	🟥
Rover	🟩	🟥	🟥	🟥	🟩	🟥	🟥	🟥	🟩	🟥
Embark	🟩	🟥	🟥	🟥	🟩	🟥	🟥	🟥	🟨 iOS only	🟥
Lyft	🟥	🟩	🟥	🟥	🟩	🟥	🟥	🟥	🟩	🟥
Uber	🟥	🟩	🟥	🟥	🟩	🟥	🟥	🟥	🟩	🟥
Roadify	🟩	🟥	🟩 Bike-sharing & Car2go	🟥	🟩	🟥	🟥	🟥	🟨 iOS only	🟥
Google Transit	🟩	🟨 GPS position only	🟥	🟥	🟩	🟥	🟥	🟩	🟨 Android only	🟥
Waze	🟥	🟩	🟥	🟥	🟩	🟥	🟥	🟥	🟩 Win phone also	🟥
Hopstop	🟩	🟥	🟥	🟥	🟩	🟥	🟥	🟩	🟨 iOS only	🟥
Zimride	🟥	🟩	🟥	🟥	🟨 Planning only	🟥	🟥	🟩	🟥	🟥
Taxi Magic	🟥	🟩	🟥	🟥	🟩	🟥	🟥	🟥	🟩	🟥
Offi	🟩	🟥	🟥	🟥	🟩	🟥	🟥	🟥	🟨 Android only	🟥
City Navigator	🟩	🟥	🟥	🟥	🟨 Prototype	🟥	🟥	🟥	🟩 HTML5	🟥
Busr	🟩	🟥	🟥	🟥	🟨 Prototype	🟨 Prototype	🟥	🟥	🟩 HTML5	🟥
The Transit App	🟩	🟥	🟥	🟥	🟩	🟥	🟥	🟥	🟩	🟥
Moovit	🟩	🟩	🟥	🟥	🟩	🟥	🟥	🟩	🟩	🟥
MATSim (open-source)	🟨 Static data only	🟨 GPS position only	🟥	🟥	🟥	🟥	🟩	🟩	🟥	🟥
M-Atlas (open-source)	🟨 Static data only	🟨 GPS position only	🟥	🟥	🟥	🟥	🟩	🟩	🟥	🟥
OneBusAway (open-source)	🟩	🟥	🟥	🟥	🟩	🟥	🟥	🟩	🟩 Win phone also	🟥
Car2go	🟥	🟥	🟩	🟥	🟩	🟥	🟥	🟩	🟩	🟥

IRMA	🟩	🟩	🟩	🟨 GPS position only – no text mining	🟩	🟩	🟩	🟩	🟩	🟨 Prototype

Remarks:
Transit data: yellow if only static data
Crowd data: data coming from users' devices, i.e. position, service sharing feeds, information feeds. Yellow if only one is present
Sensor data: data coming from external services or ad hoc sensor networks. Yellow if only one is present.
Social data: data coming from social networks, i.e. GPS position, text. Yellow if only one is retrieved
Mobile: yellow if only Android or only iOS is supported. Other platforms are a plus
Others: yellow if exceptional conditions apply