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).

Index Terms— 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. With a mobility score from 0 to 100 EU had the best regional performance with an average of 71.4 points, with seven out of the 18 analyzed cities scoring above 75 points, while Eastern/South Eastern Europe performed a regional average of 64.0 points. The estimated cost of traffic congestion to businesses and governments 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 (i.e. Arduino), 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.

Smart mobility is the target of IRMA (Integrated Real-time Mobility Assistant) project. Started in 2011 as a system to support travelers, still 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 software, and that is deployed in the cloud. Such version, named IRMA 3, 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 Art”. A systematic overview of the system 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, transport service providers. Users (including disabled people and elderly people) can benefit a real-time assistant and, of course, of an open information on transports. Municipalities and transport providers can analyze mobility status in real-time in the urban area and forecast mobility in front of unexpected and planned changes. Additionally, third parties, as tourism agencies, can use open information as a service accessible by to perform analyses and offer additional services.

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

III. STATE OF THE ART

Smart mobility services are shifting from traffic management to pursuing both the comfort for citizens and the satisfaction for authorities [11]. Unlike traditional trip planning services, smart mobility services for individuals need to support multi-modal trips and integrate green and shared transports. In addition, smart mobility services for individuals need to integrate all available internet information, namely open data, sensor data, social data and crowd data. By analyzing these 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.

To overview the state of art, we have selected a sample of 14 public mobility service systems/applications for individuals that are accessible to the public. 13 of them provide trip planning services as elementary functions; 5 offer multi-modal transport options; 3 (CityMapper, Car2go and Roadify) offer green vehicle options (e.g. charge planning for electric vehicles); 3 offer shared transport options for users (car-pooling and shared bikes), 8 deliver event alerts stemming from transport authorities or crowd-sourcing. Only 4 applications (City Navigator, Busr, MATSim and M-atlas) offer mobility analysis or mobility forecasting services. In addition, 3 applications (City Mapper, Waze and Car2go) use cloud-based architecture for data processing and analysis.

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 that are not reflecting the actual traffic conditions. Google launched Live Transit Update that uses an extended GTFS, namely 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, Boston) provide real-time sensor 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]. A demo mobile app called TrafficInfo was developed based on this architecture [15].

Currently, a promising startup called 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].

Waze uses crowd-sensing approach to get real-time positions of user’s vehicles. In order to increase user participation, local communities for drivers which enable a “collective” driving model were built. Furthermore, social interaction features and gamification techniques also increase user participation. [34] However, no public transport data are integrated. Google completed the acquisition of Waze in 2013 and started integrating crowd-sensing 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 nearby bike rental stations, available bikes and available space). Moovel and Roadify gather car sharing data (car status, i.e. 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 and only a simple demo is available [18].

B. En-route services

These scenarios are all focusing on the planning phase of a trip. With the integration of multi-modal transports and real-time data, uncertainty and disruption of transport services rise, and require en-route re-planning (i.e. compensation) to re-optimize a route during the trip [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 that automatically re-plans the trip, but without considering traffic conditions, e.g. traffic jams. In order to solve this problem, 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.
Current event alert services for smart mobility systems mainly focus on three data sources: (1) service alerts published by transport authorities and municipalities, (2) event reports submitted by citizens, (3) events detected from social networks. Let us shortly discuss them.

CityMapper, Moovit, Embark and HopStop provide service alerts published by transport authorities as users select related routes, but these alerts are static, without any analysis.

C. Information sources: crowd data and social networks

In current applications and researches, data from crowd-sensing and social-sensing 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, calculated some KPIs and found that it is such information can be useful for 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 data 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, moreover, some data are not accurate enough and some are even irrelevant.

SMARTY [22] project proposes a social-sensing approach for detecting traffic status, accidents and special events. Social-sensing module of SMARTY project uses text-mining technologies to analyze Status Update Messages (SUM). The analysis process contains SUM fetch, tokenization, stemming, Part of Speech (POS) tagging and sentences analysis. However, the actual result of social-sensing is still unknown.

Pan, Bei et al. [23] combine traditional sensing and social-sensing. Their two-step approach detects changes of travel patterns and performs text-mining on social networks; in experiments on the dataset of Beijing, it detected 86.7% of events reported by transport authorities.

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:

- Perform analytics to discover and forecast patterns about mobility of people, mobility of vehicles, environment and context;
- React in real-time to events or changing conditions;
- Integrate data sources (e.g. linked data and/or semantic web technologies);
- 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 only focuses on services for FEV users, and only sensor data (car sensor, traffic sensor and sensors from charging station) 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 multimodal 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 was developed by ATOS and deployed in Barcelona. It 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 in a contextual way for event alerts and action suggestions. However, it does not integrate social data into the platform.

Siemens mobility solution provides an Integrated Mobility Platform (IMP) which 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, on-going research projects are with a higher integration of services and data sources. Most of these integrated mobility projects support multi-modal trip planning and provide a wide range of integrated and real-time 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. We will further discuss these potential drawbacks in next chapter.

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 individual mobility, that is the focus of our IRMA project. An ideal future system should support all stages of individual mobility (Fig. 1), as (a) planning (i.e. plotting the trip end-to-end), (b) execution (i.e. en-route phase and related transactions), (c) monitoring (i.e. analysis of information on current and forecasted conditions that affect the selected itinerary, as delays, jams, disruptions, public /shared transport overcrowding), (d) control (i.e. identification and selection of an alternative route in front of condition changes, and related re-planning).

![Fig. 1. Mobility stages.](image1)

**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.

Let us consider an example of multi-layered information integration that concerns geographical information (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. In general each information layer is created by a set of web services. 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 deployed on web, smart phone, smart TV, 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].

![Fig. 2. Potential information sources in mobility.](image2)

**C. Services to user : Mobility Analyzer**

In order to exemplify IRMA 3’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. 3. Map information and mobilityt services.](image3)

Fig. 4 shows the mobility analysis layers: the first layer is the core layer that represents a static map of an urban area showing streets and transport network. The second layer highlights the routes and stops of the transports, e.g. buses, trams and subway. 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 actual position of vehicles in real time. The last layer gives the Key Performance Indicators (KPIs) on mobility.

![Fig. 4. Mobility analysis layers.](image)

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:

- 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.

- 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.

- 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? How to characterize traffic jams and congestions?

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).

![Fig. 5. During the trip - Reschedule a trip - illustrative screens.](image)

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.

![Business model diagram](image1)

**Fig. 6. Business model.**

Our business model is tech-enabled because IRMA is provided on the cloud. By providing IRMA as a service to the end user, we can benefit of 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. So, 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 crowdsourced 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**

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, clarifies the subsystems and their roles.

![Functional decomposition](image2)

**Fig. 7. Functional decomposition.**

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:

- Big data manager is the set of technologies used to store and process large data set, e.g., data warehouse, SAP Hana.
- GTFS-RT producer transforms real-time transit data from transport provider specific format to GTFS-RT format.
- Event processor deals with real-time stream processing in order to identify mobility events.

Fig. 8 shows the architecture underlying the mobility services. Next sections give details about the implementation technologies.

**A. Data availability**

Access involves a wide range of data sources:

- 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.
- Crowd data from users and/or their devices.
- 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. In Europe the term open data has been coined a few years after the Directive 2003/98/EC of the European Parliament and the Council of public sector information (PSI), which sets out the rules and practices for accessing public sector information resources for further exploitation. 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) allows public transit agencies to publish their transit data and developers to write applications that consume these data in an interoperable and standardized way. 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 classes of 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 Global Positioning System. 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 coming from TomTom API (http://developer.tomtom.com/page), in terms of:

- Traffic: a collection of calls designed to allow developers to create applications around mapping, routing, and traffic. It also allows to identify traffic condition in real-time for a specific urban area;
- Geocoding: it allows access to more functionality from the geocoding engine in order to identify correct geolocation (latitude, longitude, impact path) of traffic events;
- Traffic Cameras: it provides access to the ids and images from traffic cameras.
- Other potential sources of sensor data are relative to parking lots, and shared transport services (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 have seen a paradigm shift of the role of the user on the web, from consumer to producer of information. Due to last technological developments and their world-wide dissemination, Web 2.0 changed users’ approach to information creation and exploitation. The paradigm shift has been also a social shift. 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:

- Crowd-sourcing: travelers 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 thanks to GPS coordinates.
- Social-mining: use of location-based data from social networks is a viable way to correlate transport planning and urban activity areas. To define activity areas in a city we can model crowds by density clusters of Twitter users that activated geo-location capabilities on their mobile device, and then drill down them according to
different scales. We can also model crowds next to transit stops.

B. Data collection, storage and processing

We implemented FluenTD (http://fluentd.org/) agents by Node.js (http://nodejs.org/) on cloud for real-time data collection and filtering. Viable alternative is Python and we used it in conjunction with SAP Hana to provide a validation of social data clustering 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:

- GTFS data stored in Amazon S3 as OBJ files that describe graphs (i.e. routes) on OpenStreetMap;
- 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 transport provider specific data format (e.g. XML, JSON) to GTFS-RT;
- Sensor data about traffic coming from TomTom API in near real-time (i.e. an update every 60 seconds);
- Real-time tweets by exploiting FluentTD;
- Crowd data coming from mobile devices of travelers.

Node.js is used both for big data management and complex event processing. Data is kept raw and unstructured. In order to analyze data, we use Qubole cloud service (http://www.qubole.com/) to define models and set parameters, that is an Hadoop-as-a-service solution that implements most out of the analytics capabilities offered by Hadoop platforms (e.g. Pig, Hive, Mahout). Qubole has been designed to work on Amazon S3, so integration between the services is built-in. So, storage and analytics happen on the cloud without the need of high capability expenditures. Data connectors to Qubole 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 through web sockets.

Big Data Manager is a set of technologies exploiting Amazon Web Services and Qubole features. GTFS-RT Producer and Event Processor (implemented by Node.js) are implemented and is deployed on OpenShift cloud services.

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. The back-end is based on OpenTripPlanner (OTP) open source project, which supports GTFS and multi-modal trip planning. Additional services were also developed:

- Event awareness trip planning based on A* algorithm;
- Real-time data layer gathered from public transport authorities (LINE Pavia) and Point of Interest (POI) layer gathered from Foursquare;
- Services that support the whole life cycle of mobility issues, i.e. issue report, validation and fixing;
- Social mining and web mining services to detect mobility events in real-time;
- Data quality services (e.g. data de-duplication, reliability assessment) were implemented for validation of crowd-sourcing data and social data.

All services are deployed as Web Services on cloud. Mobility assistance services can be accessed both on web and mobile devices (e.g. Android). Mobile applications are fundamentals because the user needs to track, monitor and receive notifications about his/her trip.

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.

IRMA shall be used by a variety of mobile and fixed terminals, which include smart phone/tablet/laptop computers, smart TV to support elderly people, on board displays, etc.

1) Mobility assistance: it supports individuals during their mobility life cycle. Thus, we deliver application that can be used before the trip for planning purposes and mobile applications to execute and monitor the trip in real-time. We also deliver another application called CityFeed. It is a crowd-sourcing based collaborative application for mobility/infrastructure issue reporting and fixing. [30] CityFeed 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. These applications are available both on popular mobile platforms (i.e. Android, iOS) and on web. Citizen is the main stakeholder of mobility assistance.

2) Mobility analysis:

a) Mobility dashboard: it covers all mobility indicators and provides a territorial intelligence that integrates Geographic Information System and Business Intelligence. It consists of a web application that can be accessed by laptops,
desktop computers and mobile devices provided. Main stakeholders are municipalities and transport providers.

b) Mobility monitor: it 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 platforms that connect to the internet. Mobility monitor offer an on-the-spot service to users who do not own mobile devices and elderly people.

c) Mobility forecasting: it 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

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.

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 a step ahead 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
   a) 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.
   b) 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, influence evaluation on authors and data validation based on crowd-sourcing.

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. E.g. personal data will be pre-processed (e.g. data encryption, anonymize personal data using k-anonymity method) in local servers, and anonymized data will be transferred to public cloud for further analysis. Object matching technologies will also be used for privacy preserving in image data. E.g. human face and license numbers on images (from crowd or social data) will be detected and masked automatically.

REFERENCES


