Final Project Results Report

Deliverable ID:	**D1.3**
Dissemination Level:	**PU**
Project Acronym:	**BigData4ATM**
Grant:	**699260**
Call:	**H2020-SESAR-2015-1**
Topic:	**Sesar-02-2015 Data Science in ATM**
Consortium Coordinator:	**Nommon**
Edition date:	**27 July 2018**
Edition:	**02.00.00**
Template Edition:	**02.00.01**

Founding Members

EUROPEAN UNION EUROCONTROL

SESAR JOINT UNDERTAKING

Authoring & Approval

Authors of the document

Name/Beneficiary	Position/Title	Date
Oliva García / Nommon	Project contributor	04/05/2018
Pedro García / Nommon	Project contributor	04/05/2018

Reviewers internal to the project

Name/Beneficiary	Position/Title	Date
José Javier Ramasco / IFISC	Project contributor	07/05/2018
Riccardo Gallotti / IFISC	Project contributor	07/05/2018
Nicole Adler / HUJI	Project contributor	07/05/2018
Amir Brudner / HUJI	Project contributor	07/05/2018
Paula García / ISDEFE	Project contributor	07/05/2018
Gennady Andrienko / Fraunhofer	Project contributor	07/05/2018

Approved for submission to the SJU By - Representatives of beneficiaries involved in the project

Name/Beneficiary	Position/Title	Date
Ricardo Herranz / Nommon	Project coordinator	07/05/2018

Document History

Edition	Date	Status	Author	Justification
00.00.01	04/05/2018	Draft	Pedro García	Initial draft, used as input for the Final Review and Closeout Meeting
01.00.00	06/07/2018	Approved by Consortium	Pedro García	First version, updated with the feedback received from the Final Review and Closeout Meeting
02.00.00	27/07/2018	Approved by SJU	Pedro García	Second version, updated with the feedback received from the SJU deliverable assessment

BigData4ATM

This document is part of a project that has received funding from the SESAR Joint Undertaking under grant agreement No 699260 under European Union's Horizon 2020 research and innovation programme.

Abstract

Air transport performance objectives and decision-making processes have often overlooked the passenger perspective, mainly due to the difficulties to collect accurate, updated and reliable data on traveller needs and behaviour. The pervasive penetration of smart devices in our daily lives and the emergence of big data analytics open new opportunities to overcome this situation: for the first time, we have large-scale dynamic data allowing us to test hypotheses about passenger behaviour and intermodality. BigData4ATM is a SESAR Exploratory Research project designed to investigate how these data can be analysed and combined with more traditional demographic, economic and air transport databases to extract relevant information about passenger behaviour and use this information to inform ATM decision making processes. This report summarises the main project outcomes and describes the potential uptake of results by the SESAR programme. The report describes the applications explored within the project, discusses the way they could contribute to improving the European ATM system, and outlines the future research and innovation activities required to realise this potential.

Table of Contents

1 Executive Summary *6*

2 Project Overview *10*

2.1 Operational/Technical Context **10**

2.2 Project Scope and Objectives **11**

2.3 Work Performed **12**

2.3.1 Quality Assessment of Data Sources for ATM Socioeconomic and Behavioural Studies 12

2.3.2 Definition of Research Questions 13

2.3.3 Development of Algorithms and Methodologies for the Analysis of Passenger Behaviour from ICT-based Geolocation Data 13

2.3.4 Case studies 14

2.4 Key Project Results **15**

2.4.1 WP2: Data Identification, Collection and Assessment 15

2.4.2 WP3: Data Analysis and Predictive Modelling 16

2.4.3 WP4: Case Studies 18

2.5 Impact **23**

2.6 Technical Deliverables **23**

3 Links to SESAR Programme *25*

3.1 Contribution to the ATM Master Plan **25**

3.2 Maturity Assessment **26**

3.3 Stakeholder Feedback **29**

4 Conclusions and Lessons Learnt *30*

4.1 Conclusions **30**

4.2 Lessons Learned **31**

4.3 Recommendations for future R&D activities (Next steps) **31**

5 References *33*

5.1 Project Deliverables **33**

5.2 Project Publications **33**

Appendix A *34*

A.1 Acronyms and Terminology **34**

List of Tables

Table 1: Project Technical Deliverables 23

Table 2: Project Maturity 25

Table 3: ER Fund / AO Research Maturity Assessment 26

Table 4: Acronyms and terminology 34

List of Figures

Figure 1: BigData4ATM overall concept 11

Figure 2: Passenger centric data map 15

Figure 3: Maps with the towers involved in a Spanish air trip 16

Figure 4: Heat maps of destinations for Ibiza island 17

Figure 5: Passenger displacements at Heathrow airport from Twitter data 17

Figure 6: Travel times for each leg of the flight 19

Figure 7: Expenditures on accommodation for foreigners (left) and nationals (right) 20

Figure 8: Flight delay vs D2D travel time 20

Figure 9: K2G vs flight delay correlation 20

Figure 10: Market share per mode in each of the regions 21

Figure 11: Catchment areas of the Great London Area airports 22

1 Executive Summary

The European transport policy envisages a passenger-centric air transport system thoroughly integrated with other transport modes, with the goal of taking travellers from door to door predictably and efficiently. However, ATM operations have so far lacked a passenger-oriented perspective, with performance objectives not necessarily taking into account the ultimate consequences for the passenger. Research in this area has so far been constrained by the limited availability of behavioural data. The pervasive penetration of smart devices in our daily lives and the emergence of big data analytics open new opportunities to overcome this situation: for the first time, we have large-scale dynamic data allowing us to test hypotheses about travellers' behaviour and intermodality. The goal of BigData4ATM is to investigate how these data can be analysed and combined with more traditional demographic, economic and air transport databases to extract relevant information about passenger behaviour and use this information to inform ATM decision making. The specific objectives of the project are:

- to integrate and analyse multiple sources of passenger-centric spatio-temporal data (mobile phone records, data from geolocation apps, credit card records, etc.) with the aim of eliciting passengers' behavioural patterns;
- to develop new theoretical models translating these behavioural patterns into relevant and actionable indicators for the planning and management of the ATM system;
- to evaluate the potential applications of the new data sources, data analytics techniques and theoretical models through a number of case studies.

The project started with the identification and classification of different passenger-centric data sources, both conventional and 'Big Data' sources, analysing their strengths, limitations and potential uses. This task consisted of several sub-tasks:

- Identification of the different datasets that can be used for socioeconomic and behavioural studies in the field of ATM.
- Design of a methodology for data quality assessment.
- Assessment of data quality and identification of the main issues and uses for each data source.

The following data sources were reviewed:

- Conventional data sources: Sabre, IATA PaxIS, OAG, CODA and STATFOR public reports, EUROSTAT, national statistical offices and airport surveys.
- Big Data sources: mobile phone records, Twitter, FlightRadar24, credit card records, public transport smart card data, Google Maps APIs and TomTom Traffic Services.

For each one of these data sources, a factsheet with the following information was filled:

- General information: name, link to the source, etc.
- Abstract
- Availability: owner, access conditions, cost, etc.
- Data characteristics: size of the sample, temporal and geographical scope, etc.
- Quality issues
- Other comments

Also, for the Big Data sources, an additional analysis was conducted with the aim of characterising those aspects and particularities that are needed to consider when dealing with this kind of data, such as sample size and temporal and spatial characteristics, possible biases and representativeness of the sample.

This work produced a set of **Performance Data Factsheets** characterising each data source, which are reusable and constitute a good starting point for any future research activity that involves the use of any of the traditional or novel data sources investigated in the BigData4ATM project. In addition, the **BigData4ATM Data Repository** was also produced, which allowed the project partners to share the datasets used for the data analysis work.

In parallel to the analysis of the different data sources, a set of **research questions** was defined by the project consortium. This preliminary set of research questions was refined with the inputs from different ATM stakeholders represented in the project´s Advisory Board. The main research challenges identified can be grouped into four categories:

- **Door-to-door mobility analysis**:
    - Determine from where people go to the airport and the transport mode used.
    - Determine door-to-door travel times.
    - Develop new accessibility indicators.
    - Study new forecasting models that take into account airport catchment area.
    - Analyse competition and complementarity of air transport with ground transport.
    - Infer different passenger behavioural patterns according to their activities beyond the airport (visited places, length of the stay, transport modes used…).
    - Propose new metrics that capture how ATM delays affect total door-to-door travel time.
- **Intra-airport mobility analysis:**
    - Obtain kerb-to-gate mobility patterns.
    - Extract meaningful information related to airport processes: detect and predict flows of people and bottlenecks.
    - Optimise the allocation of airport resources according to the passenger behaviour inside the terminal.
- **Expenditure analysis:**
    - Extract expenditure patterns both inside the airport and outside the airport (e.g., tourist expenditure).
    - Explore the relationship between length of the stay at the airport and expenditure.
    - Study the impact of ATM disruptions on expenditure, how much they cost for the airport and for the tourist destinations, and whether there is an increase of the non-aeronautical revenues in airports due to an increase in passengers stay time.
- **Opinion and sentiment analysis:**
    - Study how the general public perceives the responsibility of different agents in ATM disruptions.
    - Analyse how information from official channels spreads.
    - Analyse for how long disruptions are a relevant topic in social networks.

Once the different data sources were assessed and characterised, a series of **algorithms and data analysis techniques** were developed for the extraction of new knowledge about passengers' behaviour relevant for the four categories mentioned above. This entailed:

- Development of methodologies and algorithms to infer **activity-travel patterns** at different scales. Due to their worldwide coverage, Twitter data have been used to reconstruct international passenger flows at a more aggregated level. Mobile phone records, which provide bigger samples and higher temporal granularity, have been used to reconstruct in a more detailed manner the airport access/egress legs in those countries where mobile phone data are available.
- Development of methodologies to extract **intra-airport mobility indicators**, such as heat and flow maps, from Twitter data.
- Development of methodologies to carry out **opinion and sentiment analysis** from Twitter data and extract indicators related to passenger satisfaction levels.
- Development of methodologies to extract information from credit card transactions data. This information includes **passenger expenditure** (both inside and outside the airport), comparison between different airports and seasons, and the impact of disruptions in ATM on airport non-aeronautical revenues.
- Development of statistical approaches and data fusion algorithms to **upscale the observed behaviour to the total population**.

The algorithms developed for the extraction of passenger centric mobility patterns were validated in the cases where information from other sources, such as surveys, was available. The validation exercises showed a good correlation between the information obtained from the non-conventional methods and that coming from official sources such as EUROSTATS and National Census. Mobile phone data were proven to have enough spatio-temporal granularity for the analysis of door-to-door, door-to-kerb and kerb-to-gate mobility. However, due to the fragmentation of mobile phone operators in the European context, the coverage of the mobile phone records is normally limited to a single country, restricting the analysis to one country at a time. Twitter data showed its potential for characterising transport demand at the continental level. The credit card records showed their potential to analyse how passenger expenditure patterns in and close to the airports are modified by generalised delays. Finally, it was concluded that, even when geolocated tweets are not statistically representative to produce a satisfactory analysis, semantic-sentiment analysis on the tweets' contents, once expanded to all the streaming and not only to geolocated tweets, can offer extra insights on the passenger reactions to such problematic events.

In order to test and evaluate the different methodologies developed within the BigData4ATM project, four **case studies** were defined:

- CS-1: Passenger-centric door to door travel times. The work focused on measuring the times spent in each phase of the trip (access/egress to/from the airport, stays at the airport) and explain the different patterns observed.
- CS-2: Socioeconomic impact of ATM disruptions, including the development of new indicators such as the impact of disruptions on travellers' expenditure.
- CS-3: Airports influence areas, competition/complementarity of air transport with ground transport, intermodality and competition between airports.
- CS-4: Improvement of traffic forecasting methodologies thanks to a more detailed characterisation of passenger choices.

The validation of the methodologies developed during the project by the comparison with the actual figures known by other means, as well as the results obtained in the different case studies, demonstrate that data from several sources can be blended to obtain rich and reliable information, not necessarily available otherwise, about: door-to-door travel times, broken down into the different trip segments, the impact of ATM and airport disruptions on passengers' spending behaviour and travel times, and the relevant factors affecting competition between airports and passenger modal choices.

The outcomes of the project are expected to have a positive impact on the aviation system at several levels:

- Use of new metrics and management decisions that are driven by passenger needs.
- More agile system designs, which are more resilient to challenges such as rapid changes in demand or disruptive events, thanks to the ability to evaluate ATM performance with respect to its impact on passengers.
- Seamless integration of air transport into the transport network.

The project outcomes have been positively valued by the BigData4ATM Advisory Board and other representatives of the aviation community, which makes us confident that the work done so far can advance to the next stages of the R&I cycle and constitutes the first step towards the development of marketable products and services.

2 Project Overview

2.1 Operational/Technical Context

The paramount goal of the European transport policy, as defined in the European Commission's 2011 White Paper on Transport, is to "establish a system that underpins European economic progress, enhances competitiveness and offers high quality mobility services while using resources more efficiently". Particular emphasis is put on the need for a multimodal, passenger-centric transport system, able to provide seamless door-to-door travel and facilitate better modal choices. In line with these objectives, the long-term vision for the European aviation sector outlined in the report 'Flightpath 2050 - Europe's Vision for Aviation' envisages a passenger-centric air transport system thoroughly integrated with other transport modes, with the ultimate goal of taking travellers and their baggage from door to door predictably and efficiently while enhancing passenger experience and rendering the transport system more resilient against disruptive events. The Flightpath 2050 report highlights the role of aviation as an enabler of socio-economic development, drawing attention to the key role of the ATM system in realising this vision and defining five high-level goals:

1. European citizens are able to make informed mobility choices and have affordable access to one another, taking into account: economy, speed, and tailored level of service.
2. 90% of travellers within Europe are able to complete their journey, door-to-door within 4 hours.
3. Flights arrive within 1 minute of the planned arrival time regardless of weather conditions. The transport system is resilient against disruptive events and is capable of automatically and dynamically reconfiguring the journey within the network to meet the needs of the traveller if disruption occurs.
4. An air traffic management system is in place that provides a range of services to handle at least 25 million flights a year of all types of vehicles.
5. A coherent ground infrastructure is developed including: airports, vertiports and heliports with the relevant servicing and connecting facilities, also to other modes.

In contrast with this high-level vision, ATM operations have so far lacked a passenger-oriented perspective, with performance objectives and decision criteria (e.g., flight prioritisation rules) not necessarily taking into account the ultimate consequences for the passenger. Further research is needed to provide new insights on the interactions between the ATM system and passengers' needs, choices and behaviour. However, current methods used to collect data on passengers' activities are limited in accuracy and validity: traditional methods based on observations and surveys present intrinsic limitations (e.g., incorrect and imprecise answers, dependence on the availability and willingness to answer of the interviewed persons, etc.), and they are also expensive and time-consuming. Useful data can also be collected from other sources such as air traffic databases, travel reservation systems or market intelligence data services (e.g., IATA Passenger Intelligence Services, PaxIS), but these data typically fail to capture important information, such as door-to-door origin-destination pairs and travel times. The generalised use of geolocated devices in our daily activities opens new opportunities to collect rich data and overcome many of the limitations of traditional methods. The very same ICT tools that are enabling new forms of bidirectional communication with the passenger are also making it possible to gather permanently updated information on passengers' activity and mobility patterns, with an unprecedented level of detail.

2.2 Project Scope and Objectives

The goal of BigData4ATM1 is to investigate how different passenger-centric geolocated data can be analysed and combined with more traditional demographic, economic and air transport databases to extract relevant information about passengers' behaviour, and to study how this information can be used to inform ATM decision making processes. The specific objectives of the project are:

1. develop a set of methodologies and algorithms to acquire, integrate and analyse multiple distributed sources of non-conventional spatio-temporal data coming from ICT with the aim of characterising passengers' behavioural patterns;
2. develop new theoretical models translating these behavioural patterns into relevant and actionable indicators for the planning and management of the ATM system;
3. evaluate the potential applications of the new data sources, data analytics techniques and theoretical models through a number of case studies relevant for the European air transport system.

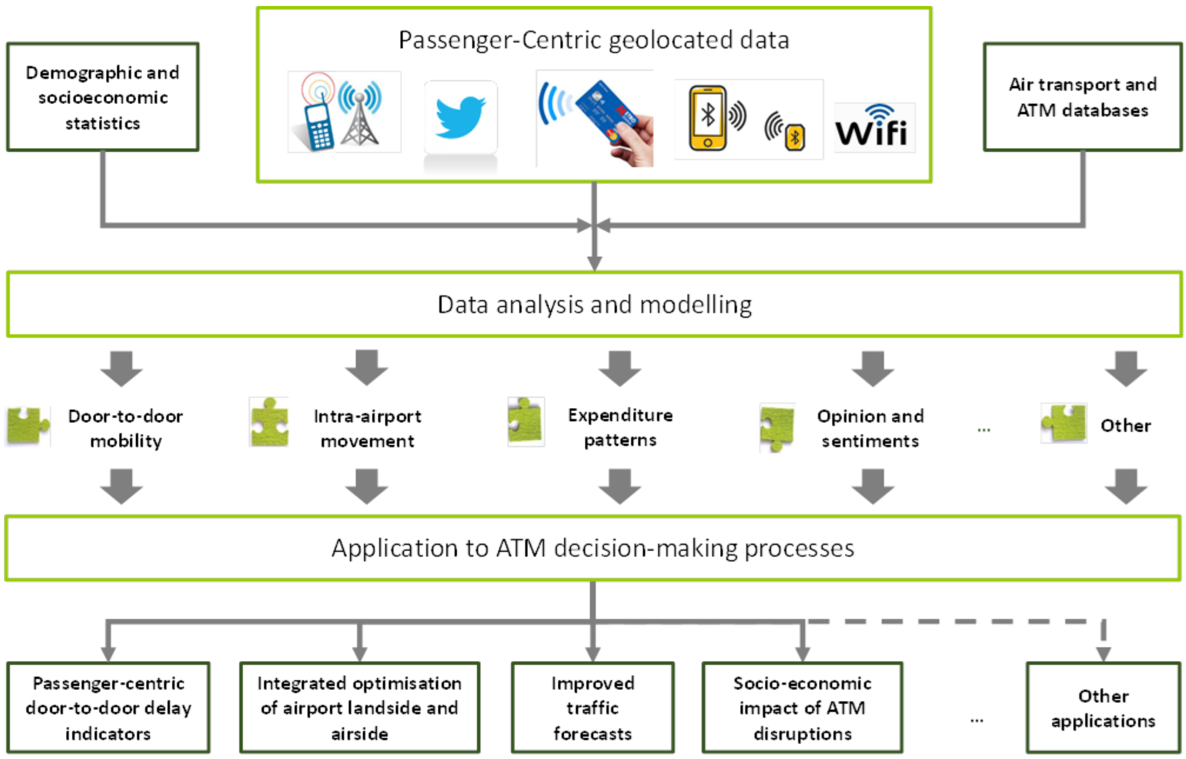


Figure 1: BigData4ATM overall concept

To achieve these goals, the project started by identifying the main gaps in traditional data sources, which resulted into a set of relevant research questions, where the use of data generated from personal mobile devices could enrich the information currently available. The preliminary research work revolved around these research questions, developing different methodologies to extract useful information from the data sources explored within the project.

Finally, the research work was structured around four different case studies (CS), each one of them corresponding to one or several of the research questions previously identified: CS-1 Passenger-centric door-to-door travel times; CS-2: Socioeconomic impact of ATM disruptions; CS-3: Airports' influence area; and CS-4: Air traffic forecasting. These case studies serve as an example of the fields where the information extracted from new, passenger-centric geolocated data sources can be of use for improving the planning and decision making processes in ATM.

2.3 Work Performed

2.3.1 Quality Assessment of Data Sources for ATM Socioeconomic and Behavioural Studies

The first task carried out within the project was to review and evaluate different datasets, including both traditional and Big Data sources, in order to characterise their strengths, limitations and potential synergies. The following data sources were reviewed:

- Conventional data sources: Sabre, IATA PaxIS, OAG, CODA and STATFOR public reports, EUROSTAT, national statistical offices and airport surveys.
- 'Big data' sources: mobile phone records, Twitter, FlightRadar24, credit card transactions, public transport smart card data, Google Maps APIs and TomTom Traffic Services[1]. For all these data sources, a detailed analysis of the quality of the information provided was carried out, assessing different aspects such as temporal and geographical scope, estimated size of the sample, availability, privacy & confidentiality issues, etc. From the analysis of traditional data sources, the following gaps and room for improvement were identified:
- There is a lack of door to door information. Although some data sources, such as surveys, provide information on places of origin and destination beyond the airport, they do not provide accurate measurements of the travel times. The door-to-kerb and kerb-to-door segments of the air trip have been very difficult to characterise from traditional data.
- There is a lack of information on airport access and egress modes, which would be of paramount importance when assessing intermodality and the integration of aviation with ground transport.
- There is also little information about passenger behaviour inside the airport. Although airports are beginning to deploy Bluetooth and WiFi sensors to monitor passenger flows and adapt their use of resources according to these flows, this is a relatively new technology and information is not yet exploited to its full potential.
- A common characteristic of many of the traditional datasets analysed is that they are too coarsely aggregated to be used for research. As an example, monthly aggregated indicators are not useful when analysing the behaviour of the ATM system on a particular day where an exogenous event impacted the ATM performance.
- Some data sources, which contain fine-grained information (often at flight level) that would be of great use for research, are of restricted access.

[1] Access to data from a mobile phone aggregator was also achieved during the last stages of the Project. This data was used for CS4, but not included in D2.1 or D3.1 due to time constraints. The analysis and methodologies followed to work with this data were included in D4.1 as an annex.

The results of the data analysis and characterisation tasks carried out during this stage of the project are described in **D2.1 Data Inventory and Quality Assessment.**

2.3.2 Definition of Research Questions

Once the review of the different data sources was carried out, a set of relevant research questions to be tackled within the next stages of the project was defined. These research questions were tightly tied to those gaps identified in traditional data sources, and were grouped into four categories:

- **Door-to-door mobility analysis:**
  - Airport catchment areas: study from where people go to the airport, explore new airport accessibility indicators, analyse competition/complementarity between airports and other transport modes, explore how this information could be used for traffic forecasting.
  - Door-to-door passenger behaviour: determine total door-to-door travel times, infer different behavioural patterns (business/leisure, accommodation…), explore how ground congestion impacts air traffic and analyse the relationship between ATM delays and door-to-door delays.
- **Intra-airport mobility analysis:** obtain kerb-to-gate mobility patterns and try to extract meaningful information related to airport processes, with the aim to detect and predict flows of people in different areas of the airport, identify bottlenecks and why they are produced, and ultimately optimise the use of airport resources for a better integration of landside and airside processes.
- **Expenditure analysis:** extract passenger expenditure patterns both inside and outside the airport and explore changes on these patterns caused by ATM disruptions.
- **Opinion and sentiment analysis:** how does the general public perceive the responsibility of different agents (ANSPs, airports, airlines)? How does information from official channels spread? How does this influence attitude towards ATM?

These research questions are detailed in **D2.1 Data Inventory and Quality Assessment**.

2.3.3 Development of Algorithms and Methodologies for the Analysis of Passenger Behaviour from ICT-based Geolocation Data

Taking the identified research questions as a starting point, a set of methodologies and algorithms were developed in order to extract relevant pieces of information from the data generated by different types of personal mobile devices. These pieces of information were grouped according to the research questions defined before:

- **Door-to-door mobility analysis:** methodologies were developed to extract mobility information from mobile phone registers, Google Maps API, Twitter geolocated data and mobile apps. With mobile phone data, the full door-to-door trip could be reconstructed, providing measurements of the airport catchment areas and the travel times for each leg of the trip. Also, methodologies to extract touristic indicators from roamers, such as length of stay, entrance airport, main destination and points of interest visited were developed. With Twitter, Google Maps and mobile apps data it was possible to develop a methodology that measured catchment areas and estimated travel times, without being restricted to a single country.

- **Intra-airport mobility analysis:** Twitter geolocated data were used to extract intra-airport mobility indicators such as flow maps and heat maps, mainly for busy airports like international hubs with several terminals.
- **Expenditure analysis:** from credit card data, a methodology to extract and characterise expenditure patterns from passengers was derived.
- **Opinion and sentiment analysis:** semantic analysis was run on the content of a set of Twitter messages on days of significant ATM disruptions, providing insights on the passenger reactions to this kind of events.

The methodologies developed were validated with official statistics, when available, which proved that the new data sources investigated by BigData4ATM can be valid sources of information, often providing a higher level of disaggregation than the one available from traditional data.

This work is documented in **D3.1 Analysis of Passenger Behaviour from ICT-based Geolocation Data.**

2.3.4 Case studies

The last part of the project comprised four case studies aimed to demonstrate the potential of the data sources analysed and the methodologies developed within the project.

The first case study, "**Passenger-centric door-to-door travel times**" analysed trip origin and destination beyond the airport, the door to door travel times and the times spent for each phase of the flight. We used mobile phone data to extract these indicators for Spanish airports, and Twitter data for the case of intra-EU flights. The results show that these methodologies provide reliable estimations of airport catchment areas and travel times, which are useful to assess the integration of air transport with the ground transport system. The results also show that the current European ATM system is far from achieving the 4 hour door-to-door goal depicted in Flightpath 2050.

The second case study, "**Socioeconomic impact of ATM disruptions**", analysed the impact of air traffic disruptive events on passenger expenditure patterns and travel times. Changes in spending patterns around the Madrid airport area during the end of May 2011, where cancellations and delays occurred, were analysed using credit card registers. The results from this analysis show that significant deviations in expenditure patterns appeared during these events. As an example, the expenditure on accommodation of foreign credit card users increased during the days of cancellations, while national credit card users' expenditure on transport increased. With regards to travel times, we analysed an episode of strikes in Barcelona during June and July 2016. The analysis revealed a correlation between average door-to-door travel times (particularly in the kerb to gate segment) and average daily delays.

The third case study, "**Airports influence areas**", studied the competition between air transport, high-speed train and road in the Madrid-Catalonia corridor by using mobile phone registers, showing how to obtain useful insights with respect to modal share and catchment areas at a lower cost than conventional methods. The results show the importance of access and egress times to airports and train stations when assessing competition between these two modes.

The fourth case study, "**Air traffic forecasting**", studied the competition between the different airports that serve the Greater London Area by using mobile apps data. In addition, it explored potential changes in the airport catchment areas due to the High-Speed 2 project, which is expected to connect London and Birmingham, bringing Birmingham airport into the London market and potentially drawing market share from Stansted, Gatwick and Luton.

The case studies are documented in **D4.1 Applications of Passenger-Centric Geolocation Data to the Planning and Management of the ATM System: Case Studies**.

2.4 Key Project Results

2.4.1 WP2: Data Identification, Collection and Assessment

In this work package, the different datasets that were used within the project were gathered and analysed. In addition, the set of research questions that would be addressed by combining these datasets were defined. In particular, the following results were produced:

- A set of factsheets that contained the characterisation of the main data sources useful for ATM socioeconomic and behavioural studies, assessing their strengths and weaknesses. Traditional sources studied were Sabre, IATA PaxIS, OAG, CODA and STATFOR public reports, EUROSTAT, national statistical offices and airport surveys. Novel sources included mobile phone records, Twitter, FlightRadar24, credit card transactions, public transport smart card data, Google Maps APIs, TomTom Traffic Services and mobile apps data. These factsheets are reusable and constitute a valuable resource for any future research activity that involves the use of any of the traditional or novel data sources investigated in the BigData4ATM project.
- A "passenger-centric" data map was created, linking each of the novel data sources with the domains where they can provide insights.

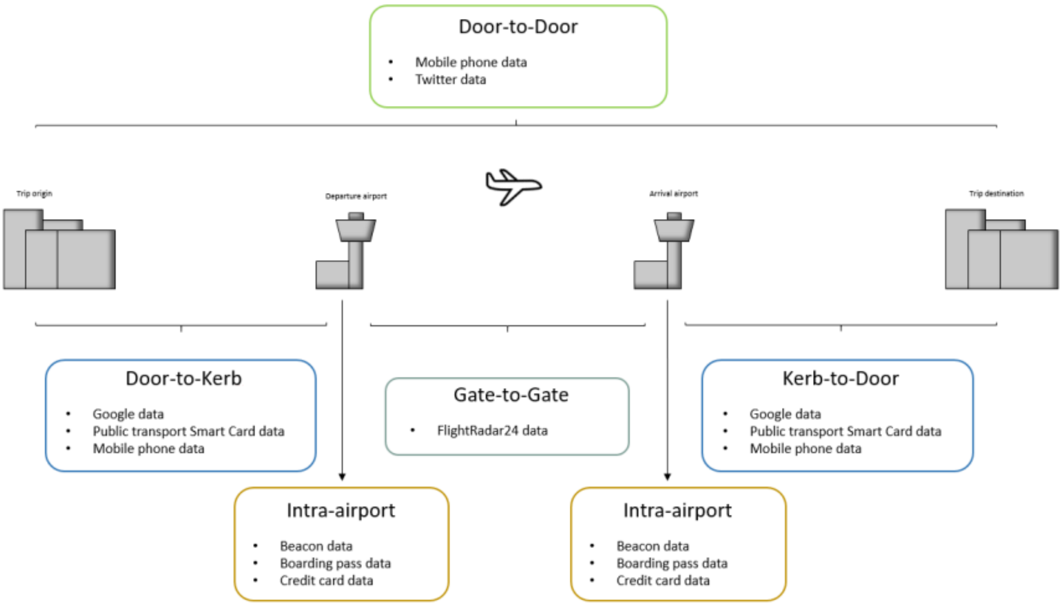


Figure 2: Passenger centric data map

- From the analysis of traditional data sources, it was concluded that the door to kerb and kerb to door segments of a trip, that is, those legs beyond the airports, are often overlooked. Exploring how the knowledge on these legs could be increased by the use of novel data sources would become one of the cornerstones of the project.
- A set of relevant questions related with passenger mobility and behavioural patterns was derived from the characterisation of the data sources and the consultation with several

stakeholders. These research questions not only provided the guide for the analysis carried out within the project, but it also constitutes a preliminary agenda for future passenger-centric research.

2.4.2 WP3: Data Analysis and Predictive Modelling

The different data sources gathered by the project were analysed and combined to address some of the research questions previously identified. The following methodologies were developed:

- The algorithms present in the literature for extracting mobility information from mobile phone records were adapted to match the requirements of air transport. In particular, this data source has traditionally been used for urban mobility projects, and several adjustments were needed to adequately capture air trips. An example of a trip detected with this new methodology can be observed in Figure 3.

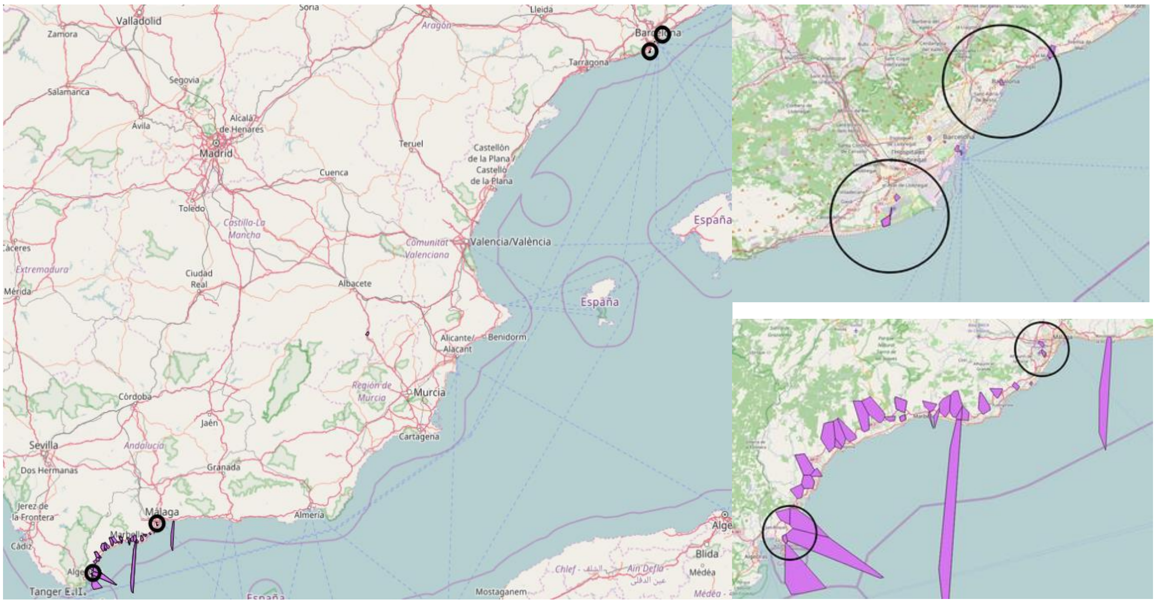


Figure 3: Maps with the towers involved in a Spanish air trip

- A methodology to extract touristic information from the mobile phone registers produced by roamers was developed. The main problem here was to adapt the existing algorithms to a set of the population with a lower register generation rate. In addition, the sampling framework for these users was also different than for national users, being necessary to use official touristic statistic to upscale the observed sample. As an example of the information produced, heat maps of destinations for tourists and locals for Ibiza are shown in Figure 4.

Figure 4: Heat maps of destinations for Ibiza island

- Algorithms were developed that, from geolocated Tweets, extracted origins and destinations of air trips and, for the case of big airports, passenger heat maps and flows between terminals. An example of the information produced is shown in Figure 5.

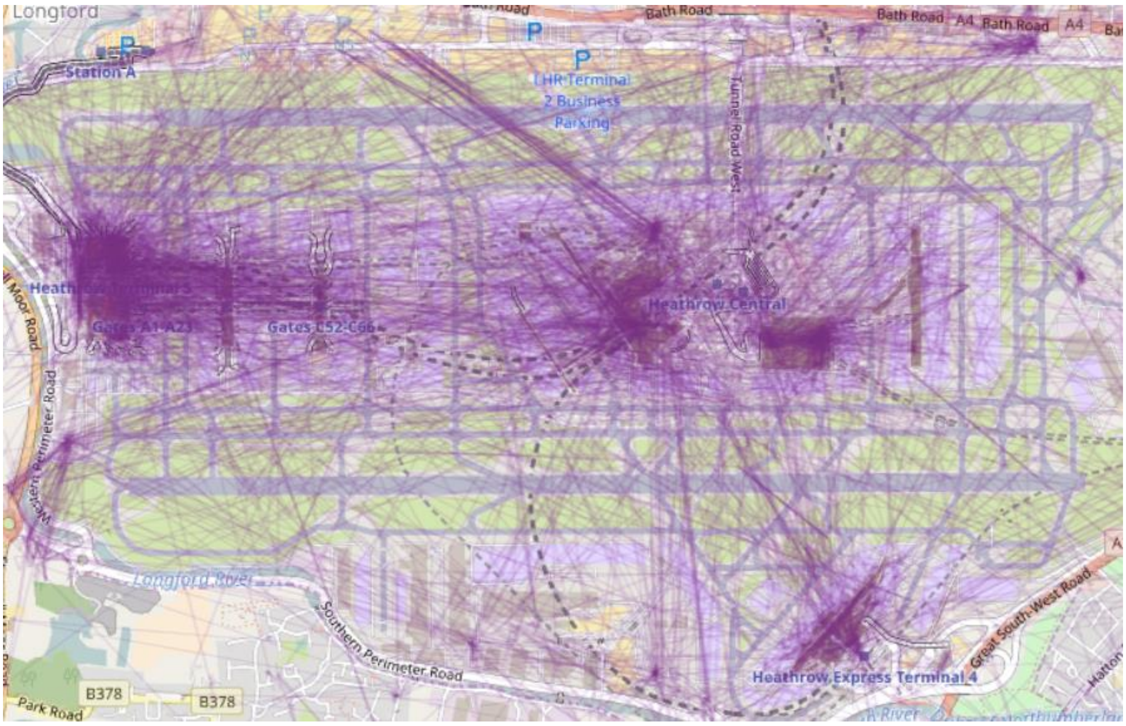


Figure 5: Passenger displacements at Heathrow airport from Twitter data

- Methodologies to filter out air transport passengers from the credit card sample of users and to scale them to the total population were developed. Through these methodologies, additional behavioural information could be derived for tourists, including expenditure patterns (what they buy, where do they buy, etc.) but also length of stay and place of accommodation.

2.4.3 WP4: Case Studies

The data collected and methodologies developed in the previous stages in the BigData4ATM project were combined in order to provide actionable information for ATM decision making processes in four specific use cases.

Case Study 1: Passenger-centric door to door travel times

Mobile phone records and geolocated Twitter data were used to extract information on door to door travel times, broken down into the different legs of the trip (Figure 6). From the results obtained, we can extract the following conclusions:

- Intra EU travel times are far from the 4h door to door target. In the case of Madrid airport, for peninsular flights (around 1 hour of duration), total door to door goes up to 5.5 hours on average, while for the flights from the Canary Islands (flights of around 2.5 hours of duration) the total door to door increases up to 7.5 hours on average.
- Measured travel times indicate that passengers spend a significant amount of time at airports, particularly at the origin airport, where on average passengers stay for around 2 hours. This result suggests that, in order to reduce total door to door travel times, passengers should spend less time at airports, which may have an undesired impact on non-aeronautical airport revenues.

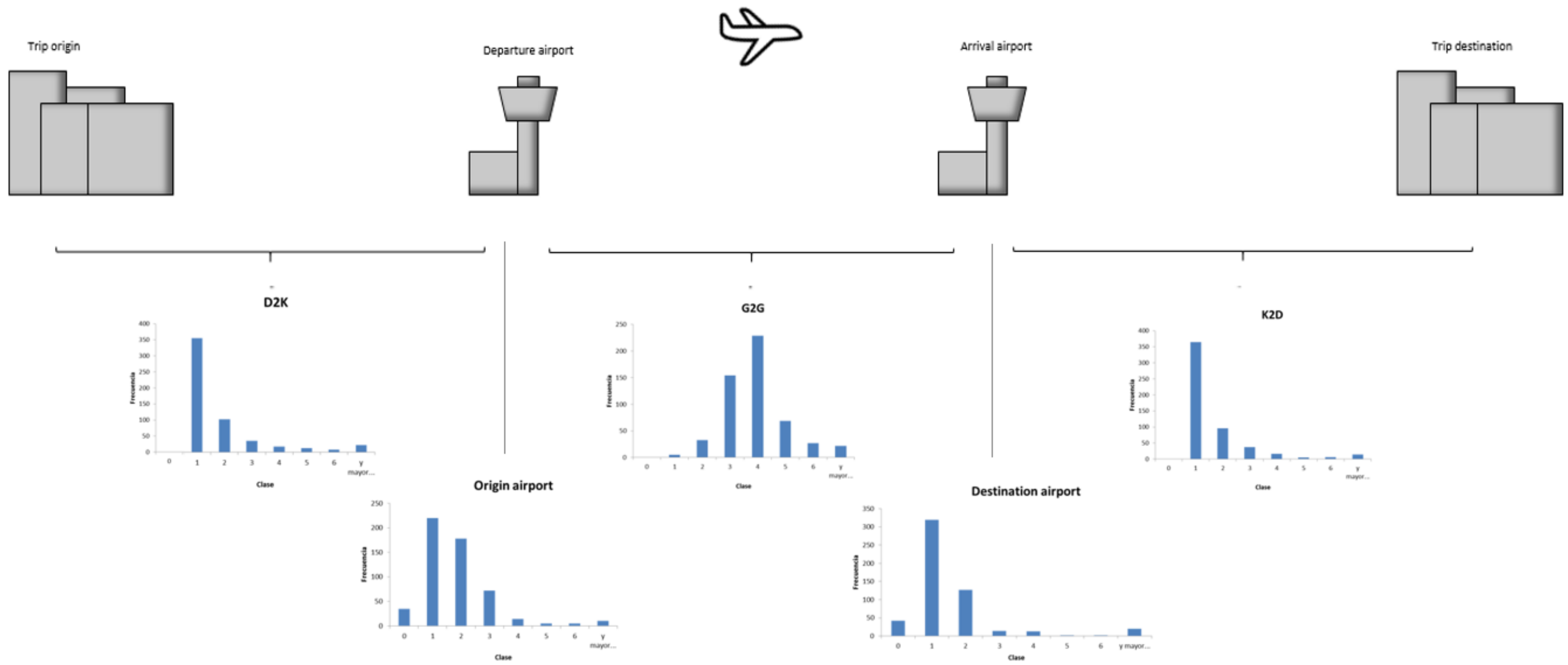


Figure 6: Travel times for each leg of the flight

Case Study 2: Socioeconomic impact of ATM disruptions

Credit card data and mobile phone registers were used to assess the impact of disruptive events.

- Expenditure analysis shows that, as expected, foreign tourists are affected differently from national travellers: foreign need to look for accommodation (Figure 7), while in the case of nationals, they might feel more comfortable changing their mode of transport.

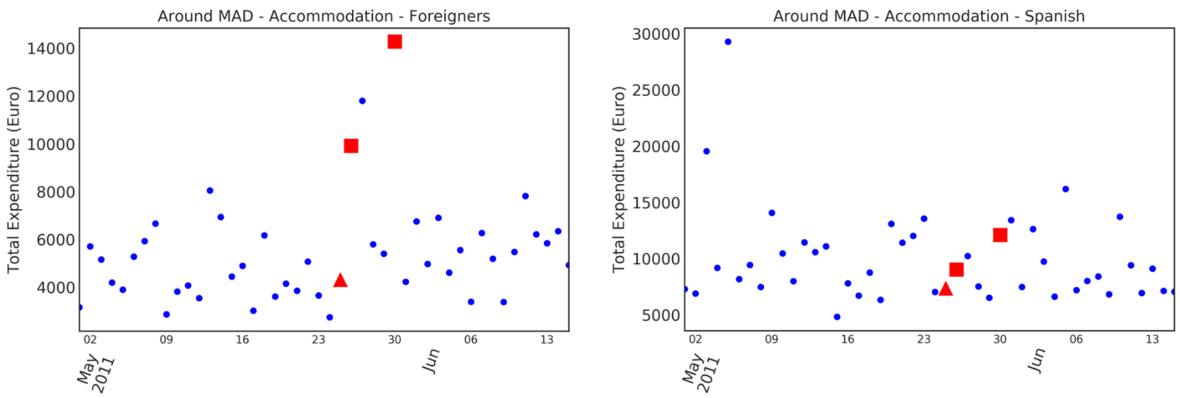


Figure 7: Expenditures on accommodation for foreigners (left) and nationals (right)

- A good correlation was found between door to door travel times and flight delays (Figure 8), and also (as expected) between kerb to gate travel times and flight delays (Figure 9).

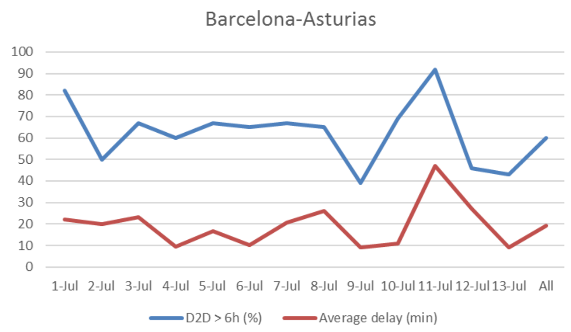


Figure 8: Flight delay vs D2D travel time

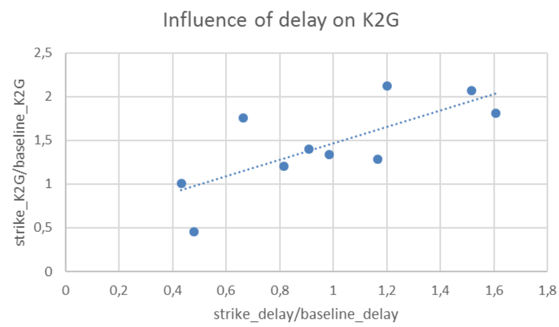
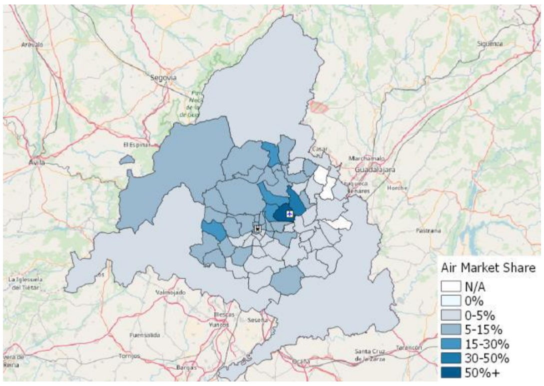


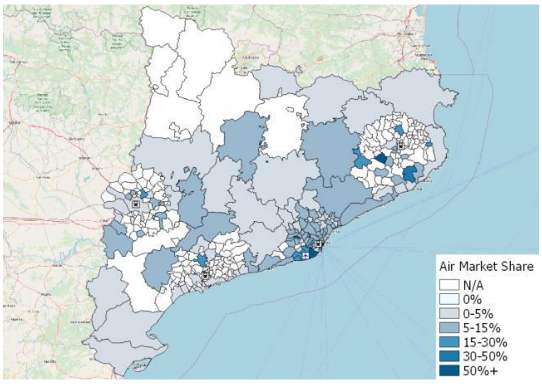
Figure 9: K2G vs flight delay correlation

Case Study 3: Airports influence areas

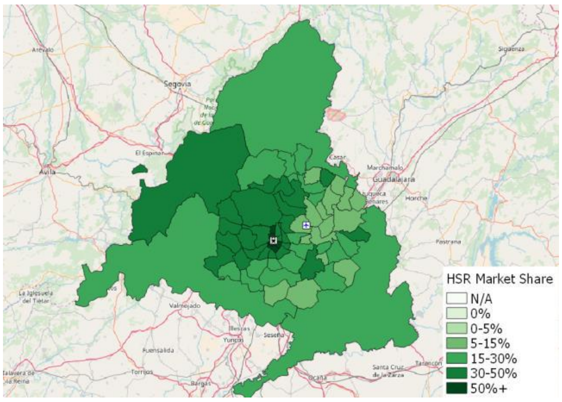
Mobile phone records were used to obtain the modal share of the Madrid-Catalonia corridor, determining the competition between air transport, high speed train and road. The catchment areas for each mode were derived, showing that the better the accessibility to an airport/train station, the more attractive these transport modes are (Figure 10). Through a multinomial logit model, the influence of access and egress times to airports/train stations was quantified. Results show that the longer the trip, the lower is the probability that the road network is chosen. Also, the results show that airline passengers are more sensitive to accessibility to the airport than train passengers.



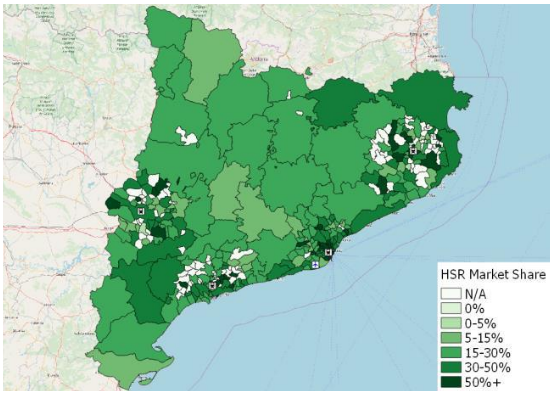
Aviation market share in Madrid region



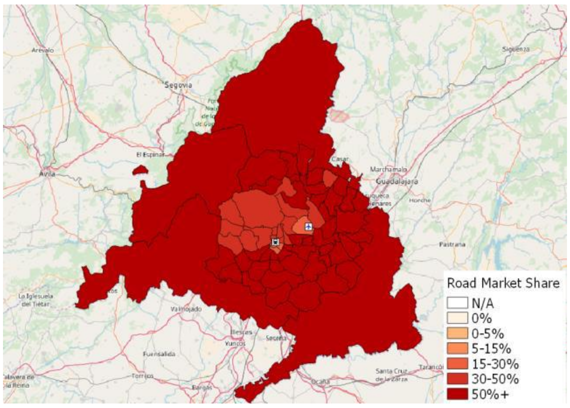
Aviation market share in Catalonia region



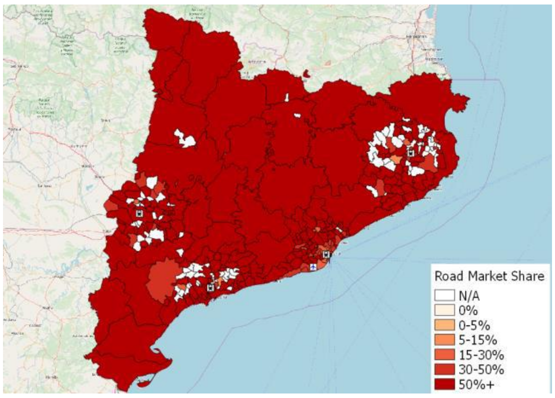
High speed train market share in Madrid region



High speed train market share in Catalonia region



Road market share in Madrid region



Road market share in Madrid region

Figure 10: Market share per mode in each of the regions

Case Study 4: Air traffic forecasting

Data from a mobile phone app was used to determine if there is competition between the airports in Greater London Area, and in affirmative case, to identify the factors that affect this competition. Also, the impact of the opening of the London-Birmingham High Speed Train connection was explored. Results show that Heathrow attracts demand from all England and Wales, while the rest of London airports have catchment areas much narrower. Therefore, it seems that these airports avoid competition (Figure 11). Based on a set of logistic regressions, the results show that a Birmingham-London travel time of 39 minutes by rail would draw market share from Stansted and possibly Gatwick and Luton, increasing Birmingham airport traffic by 30%.

Figure 11: Catchment areas of the Great London Area airports

2.5 Impact

The new data sources explored and the methodologies and algorithms developed to extract useful information from these sources will lead to **an improvement of the quality & availability of knowledge for decision making**. As shown within the project, these methodologies can be applied to study both normal conditions and unexpected events that caused disruptions in the ATM system. This capability can be further explored by developing recommendations and decision support tools contributing to a **more agile and robust ATM system**. The project has also studied the competition and complementarity between air transport and high-speed train, showing that, although the opening of high speed rail lines has a significant impact on air transport, there is room for cooperation between these modes. This goes in the direction of **enhancing intermodality and the integration of air transport into the ground transport network, exchanging passengers seamlessly**. In addition, the project has developed methodologies and metrics regarding door-to-door travel times in Europe, which will contribute to achieving **performance benefit across the entire ATM system**: now that there is a methodology to measure D2D, it is possible to assess the level of accomplishment of several performance objectives and the impact of policies and measures that aim to reduce these travel times. In conclusion, BigData4ATM has contributed to **promoting a passenger-centric approach to research in ATM**, opening promising research avenues related to this field.

2.6 Technical Deliverables

Table 1: Project Technical Deliverables

Reference	Title	Delivery Date[2]	Dissemination Level[3]
Description			
D2.1	**Inventory and Quality Assessment of Data Sources for ATM Socioeconomic and Behavioural Studies**	**15/11/2016**	**Public**
This document presents an analysis of the data sources currently available for socioeconomic and behavioural research in ATM, including both traditional air transport and demographic data sources and new, unconventional passenger data generated by mobile devices. An analysis of these data sources has been carried out, identifying gaps in traditional data sources and potential synergies with new 'Big Data' sources. Finally, the research questions that would be tackled during next stages of the project were outlined.			
D3.1	**Analysis of Passenger Behaviour from ICT-based Geolocation Data**	**02/02/2018**	**Public**
This document presents the methodology and the algorithms developed to extract relevant information about passenger behaviour from the data sources described in deliverable D2.1. The results obtained with the proposed methodology are validated by comparing them with those obtained through alternative, more traditional data sources. The advantages and limitations of the information extracted from new, big data sources are discussed with a view to inform their use in the case studies to be developed in WP4.			

[2] Delivery data of latest edition

[3] Public or Confidential

Reference	Title	Delivery Date[2]	Dissemination Level[3]
Description			
D4.1	Applications of Passenger-Centric Geolocation Data to the Planning and Management of the ATM System: Case Studies	23/04/2018	Public
This document describes the work performed in BigData4ATM to bring the methodologies and the algorithms developed within the project to the point where they can be used to provide actionable information for decision making. The added value of the data and methodologies investigated by the project is demonstrated through their application in four Case Studies related with door to door travel times, socio economic impact of ATM disruptions, airport catchment areas and air traffic forecasting.			

3 Links to SESAR Programme

3.1 Contribution to the ATM Master Plan

BigData4ATM aims to provide new tools for better accounting for the passenger perspective in ATM decision making. The main SESAR Solutions that could benefit from the project results are the following:

Table 2: Project Maturity

Code	Name	Project contribution	Maturity at project start	Maturity at project end
New enabler	Passenger centric metrics for airport Performance Planning and Monitoring	The methodologies developed within BigDat4ATM to measure door to door travel times, assess airport catchment areas and evaluate the competition/cooperation between air transport and high speed train would be of great value for Total Airport Management (TAM). This suggested Enabler supports the Operational Improvement AO-0802-B by making it possible to use passenger centric metrics of landside activities inside and outside the airport for improved airport performance planning and monitoring	TRL-0	TRL-1
Transversal	Performance Management	The methodologies developed within BigData4ATM can be applied for real world performance measurement. In the case studies developed within the project, it has been shown that the information that can be obtained from passenger mobile devices can be used to assess current ATM performance and to support the long-term steering and development of the Air Transport System.	TRL-0	TRL-1

3.2 Maturity Assessment

Table 3: ER Fund / AO Research Maturity Assessment

ID	Criteria	Satisfaction	Rationale - Link to deliverables - Comments
TRL-1.1	Has the ATM problem/challenge/need(s) that innovation would contribute to solve been identified? Where does the problem lie?	Achieved	In D2.1, we have identified a lack of passenger-related information in the traditional sources used for socioeconomic and behavioural studies in ATM.
TRL-1.2	Has the ATM problem/challenge/need(s) been quantified?	Not applicable	In the Final Review it was considered as not applicable for this kind of ER.
TRL-1.3	Are potential weaknesses and constraints identified related to the exploratory topic/solution under research? - The problem/challenge/need under research may be bound by certain constraints, such as time, geographical location, environment, cost of solutions or others.	Achieved	Constraints regarding data availability, access to restricted data sources, geographically restricted data and spatio-temporal aggregation have been identified.
TRL-1.4	Has the concept/technology under research defined, described, analysed and reported?	Achieved	BigData4ATM has developed several methodologies to complement the traditional data sources in terms of passenger-centric information. In particular, information about door to door travel times, airport catchment areas, intermodality and airport accessibility are described in D3.1
TRL-1.5	Do fundamental research results show contribution to the Programme strategic objectives e.g. performance ambitions identified at the ATM MP Level?	Achieved	More efficient and informed decision making processes enabled by new passenger-centric performance metrics, more efficient ATM and airport resource planning, in particular when referred to D2D measurement technologies.

ID	Criteria	Satisfaction	Rationale - Link to deliverables - Comments
TRL-1.6	Do the obtained results from the fundamental research activities suggest innovative solutions/concepts/ capabilities? - What are these new capabilities? - Can they be technically implemented?	Partially achieved	New solutions are outlined in D1.3. These include new tools for airport planning and management, as well as new solutions for the exchange of passenger. In particular, methodologies to assess the intermodality are suggested.
TRL-1.7	Are physical laws and assumptions used in the innovative concept/technology defined?	Achieved	Due to the broad and exploratory nature of the different case studies of the project, several assumptions needed to be made. The assumptions for the models and algorithms developed by the project are described in D3.1 and D4.1. As examples, we can mention assumptions related to individual behaviour to determine home location from mobile phone data, or whether a user has performed an air trip through Twitter data were made.
TRL-1.8	Have the potential strengths and benefits identified? Have the potential limitations and disbenefits identified? - Qualitative assessment on potential benefits/limitations. This will help orientate future validation activities. It may be that quantitative information already exists, in which case it should be used if possible.	Achieved	Strengths and limitations are described in D2.1, D3.1 and D4.1, and summarised in D1.3. The main strength of these new data sources is that they can provide useful information to overcome the lack of passenger-centric information in ATM, and the main limitation is that, as these data were not designed for this kind of studies, a careful and rigorous approach needs to be made to not ending with a biased, useless sample of users.

ID	Criteria	Satisfaction	Rationale - Link to deliverables - Comments
TRL-1.9	Have Initial scientific observations been reported in technical reports (or journals/conference papers)?	Achieved	In addition to project deliverables, several journal and conference papers have been produced.
TRL-1.10	Have the research hypothesis been formulated and documented?	Achieved	The research hypotheses are described in D3.1 and D4.1. These hypotheses depend on the data source used and the objective of the case study. As an example, we can mention the hypothesis related to tourist/excursionist determination in roamers mobile phone data.
TRL-1.11	Is there further scientific research possible and necessary in the future?	Achieved	Future research is outlined in D1.3. We consider particularly interesting, deploying and testing some of the methodologies derived from BigData4ATM in an airport to extract passengers' mobility patterns to spot/predict capacity-demand imbalances, enabling a rapid and efficient adaptation of the airport resources (e.g., check-in counters, security controls) according to changes in demand.
TRL-1.12	Are stakeholder's interested about the technology (customer, funding source, etc.)?	Achieved	Significant interest has been detected through different stakeholder consultation mechanisms, especially from airports. For more information regarding stakeholder's feedback, refer to D1.3 section 3.3.

3.3 Stakeholder Feedback

In general, the project has delivered interesting results that have been positively valued amongst industrial and academic stakeholders.

The most active and interested stakeholders were airports, as some of the outcomes of the project can provide valuable information about the door-to-kerb and kerb-to-door segments, for which little information exist, despite their importance for airports. Increasing the availability of information regarding the passenger will enhance airport management, enabling an efficient and flexible use of resources. Another interesting piece of feedback received was that not all stakeholders benefit equally from the same performance goals, e.g., it seems that the area with more room for improvement in the D2D trip in terms of travel time is the airport, due to the great time buffers taken by passengers to mitigate any unexpected event. However, airports benefit economically from these idling passengers at their facilities, as they typically involve some sort of expenditure.

Some of the results were considered by the stakeholders as mature enough to be included into a more advanced phase of research, by working with an airport to develop a pilot study in an operational environment.

Finally, the work carried out within the project has allowed us to identify several commonalities and synergies with other initiatives, such as:

- Work carried out at STATFOR to develop a connectivity map of Europe by combining airport ground accessibility with flight schedules.
- Work carried out in the U.S. (Georgia Tech), where they are using mobile phone data to assess door to door travel times.
- The ESSnet BigData project, which aims at the integration of big data in the regular production of official statistics.

4 Conclusions and Lessons Learnt

4.1 Conclusions

BigData4ATM has demonstrated the potential of data coming from smart personal devices to support a passenger-oriented perspective when designing ATM performance objectives. The project has shown the usefulness of these data for the definition of passenger centric indicators able to capture the ultimate consequences of ATM performance on the passengers.

The new methodologies and algorithms for the extraction of passenger's behavioural patterns from different sources of passenger centric data have been demonstrated through a series of case studies. In particular the project has proven the usefulness of such passenger centric data to understand the interaction between air transport and ground transport, and the role of air transport in the full door to door passenger journey. This understanding may help to develop a more integrated approach for the planning and management of the European transport system.

The main outputs of the project are:

- a detailed review of available databases relevant for the analysis of passenger's behaviour and its relation with ATM performance, with particular attention to passenger centric data sources;
- a thorough analysis of the strengths and limitations of the identified data sources;
- a set of algorithms for the extraction of passengers mobility and expenditure patterns from geolocated passenger centric data and a methodology to upscale the sample data to the total population;
- reliable indicators for the door to door travel times, differentiating the time spent in each phase of the trip;
- enhanced models for air traffic demand forecasting;
- enhanced models for airport's competition and competition between air transport and other modes of transport.

The outcomes of this work are expected to have a positive impact on the ATM system at several levels, such as:

- The design of more agile ATM system designs that are more resilient to challenges such as rapid changes in demand or disruptive events, thanks to the ability to evaluate ATM performance through the impact on passengers.
- A seamless integration of ATM into the transport network.
- The use of new metrics and management decisions driven by passenger needs.

Despite being of an eminently exploratory nature, the research activities proposed by BigData4ATM entail a significant innovation potential and can open new market opportunities in several areas:

- New analytics products and services replacing or complementing the traditional methods used to gather information on passenger behaviour. These applications go beyond ATM, and can be of interest for all the stakeholders of the aviation sector.
- Innovative demand forecasting methodologies and tools.
- New decision-support tools to better respond to passenger needs and expectations. Particularly interesting are the opportunities to develop decision support tools for airports.

4.2 Lessons Learned

The work conducted during this project allowed consortium members to:

- acquire a deeper understanding of the different datasets available for socioeconomic and behavioural studies in ATM, their gaps and the potential of Big Data sources to overcome them. However, there exists some key aspects that need to be carefully taken into account when working with this kind of data, such as the need for statistical and mathematical rigour, the fact that each data source has its own range of applicability, thus leading to the need of data fusion, and the importance of validation exercises with ground truth data;
- strengthen their position regarding the application of data analytics to aviation and ATM. For example, Nommon is currently working with the Spanish Ministry of Transport on a pilot project aimed to explore how some of the algorithms developed in the frame of BigData4ATM can be used for the analysis travel demand along the main long-distance transport corridors in Spain.

Additionally, from a project management perspective there are also several lessons learnt that might of help for subsequent R&D activities:

- Ethical issues shall be taken into account from the very beginning of the project. This is particularly relevant in the context of the new European General Data Protection Regulation. If the research lines identified within the project are going to continue into industrial and applied research, ethical requirements may be more severe, so it will be essential to allocate resources to fulfil the relevant requirements from the competent European and national authorities
- Administrative and financial reporting could be streamlined, which would release resources (on both SJU and project team sides) for scientific-technical audit and strategic planning.
- There is great benefit from strong stakeholder involvement, either led by the project (Stakeholder Workshops), facilitated by the SJU (e.g., SJU-ACI Workshop) or through participation in scientific congresses and forums.
- Project duration of 24 months is a limiting factor to develop the full potential of ER projects, which is aggravated by the reporting requirements and the need to prepare the Final Report and hold the Final Review before the project end.

The project has also allowed the identification of future activities required to further develop the project results, which are outlined in section 4.3 below.

4.3 Recommendations for future R&D activities (Next steps)

The results obtained in BigData4ATM and the feedback provided by different stakeholders gathered in the workshops organised by the project point towards several research question to be addressed in future projects:

- Extension of the analysis conducted with mobile phone data to the rest of Europe, in order to assess door-to-door travel times and origin/destinations at European level.
- Enrichment of the passenger's data with socio-demographic information such as age, gender and income. The use of other socio-demographic data sources, together with data fusion and machine learning techniques, could allow a better characterisation of passenger's profile, which would help to improve mode choice and airport selection models.

- Enrichment of mobility patterns by integrating other data sources not included in this project, such as data from airports/airlines apps to better characterise the intra airport behaviour.
- Evaluation of the newly developed solutions in a real life environment and development of predictive models for decision making in collaboration with an airport. The new decision support tools would rely on more accurate information about passengers' mobility patterns when arriving/exiting the airport and inside the airport to spot/predict capacity-demand imbalances, enabling a rapid and efficient adaptation of the airport resources (e.g., check-in counters, security controls) according to changes in demand.

5 References

5.1 Project Deliverables

- D1.1 Project Management Plan 01.00.00 Nov 2016
- D1.2 Data Management Plan 01.00.00 Nov 2016
- D2.1 Inventory and Quality Assessment of Data Sources for ATM Socioeconomic and Behavioural Studies 01.00.00 Nov 2016
- D3.1 Analysis of Passenger Behaviour from ICT-based Geolocation Data 01.01.00 Feb 2018
- D4.1 Applications of Passenger-Centric Geolocation Data to the Planning and Management of the ATM System: Case Studies 02.00.00 Jul 2018
- D5.1 Project Website
- D6.1 H - Requirement No. 1 00.01.00 Jun 2016
- D6.2 POPD - Requirement No. 2 00.01.00 Dec 2016
- D6.3 POPD - Requirement No. 3 00.01.00 Jul 2016
- D6.4 POPD - Requirement No. 4 00.02.00 Nov 2017
- D6.5 DU - Requirement No. 5 00.01.00 Jun 2016
- D6.6 OEI - Requirement No. 6 00.01.00 Jun 2016
- D1.3 Project Results Final Report 02.00.00 Jul 2018

All public deliverables are available for download from the Publications section of the project website: https://www.bigdata4atm.eu/publications

5.2 Project Publications

- P. García-Albertos, M. Picornell, M. Salas-Olmedo, J.Gutiérrez (2018) Exploring the potential of mobile phone records and online route planners for dynamic accessibility analysis.Transportation Research Part A https://doi.org/10.1016/j.tra.2018.02.008
- P. García-Albertos, O.G. Cantú Ros, R. Herranz (2018) “Analyzing Door to Door Travel Times Through Mobile Phone Data”, in ICRAT 2018 proceedings, best paper award on the economics & policy track.
- Riccardo Gallotti, Rémi Louf, Jean-Marc Luck and Marc Barthelemy (2018) “Tracking Random Walks”, Journal of The Royal Society Interface, 15(139), 20170776.
- P. García-Albertos, O.G. Cantú Ros, C. Ciruelos and R.Herranz (2017) “Understanding Door-to-Door Travel Times from Opportunistically Collected Mobile Phone Records: A Case Study of Spanish Airports”, in D. Schaefer (Ed.) Proceedings of the SESAR Innovation Days 2017, EUROCONTROL.
- Gennady Andrienko, Natalia Andrienko, Wei Chen, Ross Maciejewski and Ye Zhao (2017) “Visual Analytics of Mobility and Transportation: State of the Art and Further Research Directions”, IEEE Transactions on Intelligent Transportation Systems, Volume: PP, Issue: 99, pp. 1-18.
- R. Gallotti, M. Fuster and J.J. Ramasco (2017) “New Data Sources to Study Airport Competition”, in D. Schaefer (Ed.) Proceedings of the SESAR Innovation Days 2017, EUROCONTROL.
- P. García-Albertos, J.J. Ramasco, G. Andrienko, N. Adler, C. Ciruelos and R. Herranz (2016) “Big Data Analytics for a Passenger-Centric ATM System: A Case Study of Door-to-Door Intermodal Passenger Journey Inferred from Mobile Phone Data”, in D. Schaefer (Ed.) Proceedings of the SESAR Innovation Days 2016, EUROCONTROL.

Appendix A

A.1 Acronyms and Terminology

Term	Definition
ANSP	Air Navigation Service Provider
ATM	Air Traffic Management
CODA	Central Office for Delay Analysis
IATA	International Air Transport Association
ICT	Information and Communication Technologies
SESAR	Single European Sky ATM Research Programme
SJU	SESAR Joint Undertaking (Agency of the European Commission)
STATFOR	Statistics and Forecasts
SWIM	System-Wide Information Management

Table 4: Acronyms and terminology