Resilience2050.eu

New design principles fostering safety, agility and resilience for ATM

Grant agreement no: 314087

Theme AAT.2012.6.2-4. Building agility and resilience of the ATM system beyond SESAR

Funding Scheme: Collaborative Project (small or medium-scale focused research project)

D2.1 Description of the integrated database schema

Revision: Final Version
Date: 31/01/2013
Status: Final document
Dissemination Level: CO

<table>
<thead>
<tr>
<th>Documentation</th>
<th>Organisation</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prepared</td>
<td>The Innaxis Research Foundation and Institute</td>
<td>Massimiliano Zanin, Hector Ureta</td>
</tr>
<tr>
<td>Reviewed</td>
<td>Resilience2050 Consortium</td>
<td></td>
</tr>
<tr>
<td>Approved</td>
<td>The Innaxis Research Foundation and Institute</td>
<td>David Perez</td>
</tr>
</tbody>
</table>
## RECORD OF REVISIONS

<table>
<thead>
<tr>
<th>Revision</th>
<th>Date</th>
<th>Reason for Review</th>
<th>Modified sections/page</th>
</tr>
</thead>
<tbody>
<tr>
<td>v0.01</td>
<td>10/01/2013</td>
<td>Creation of the document</td>
<td>All</td>
</tr>
<tr>
<td>v0.02</td>
<td>20/01/2013</td>
<td>Version for review by partners</td>
<td>All</td>
</tr>
<tr>
<td>v0.03</td>
<td>30/01/2013</td>
<td>Version reviewed, ready for final approval</td>
<td>All</td>
</tr>
<tr>
<td>v0.04</td>
<td>31/01/2013</td>
<td>Final version</td>
<td>All</td>
</tr>
</tbody>
</table>
1 INTRODUCTION
  1.1 ACRONYMS AND TERMINOLOGY
2 WP2 IN RESILIENCE2050
3 D 2.1 IN WP 2
4 OBJECTIVES, CHALLENGES AND REQUIREMENTS OF D 2.1
  4.1 OBJECTIVES AND CHALLENGES
  4.2 REQUIREMENTS OF D 2.1
    Scenario 01 – Delays due to changes in the airport configuration
    Scenario 02 – Delay generation due to staffing problems at one airport
    Scenario 03 – Restrictions in ATC capacity
    Scenario 04 – Trajectories modifications due to adverse weather
5 SPECIFICATION OF DATA SOURCES
  5.1 PRISME DATA
  5.1.1 PRISME
  5.1.2 The ALL-FT+ data set
  5.2 METAR DATA
  5.3 ATATURK WEATHER DATA
  5.4 ACC HEADLINES
  5.5 ENVIRONMENTAL DATA SET
  5.6 EUROCONTROL ATFM – NOP
6 ANNEXES
  6.1 DATA CONFIDENTIALITY AGREEMENT ON THE USE OF ALL FT+ DATASET

---

TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>INTRODUCTION</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>1.1 ACRONYMS AND TERMINOLOGY</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>WP2 IN RESILIENCE2050</td>
<td>7</td>
</tr>
<tr>
<td>3</td>
<td>D 2.1 IN WP 2</td>
<td>8</td>
</tr>
<tr>
<td>4</td>
<td>OBJECTIVES, CHALLENGES AND REQUIREMENTS OF D 2.1</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>4.1 OBJECTIVES AND CHALLENGES</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>4.2 REQUIREMENTS OF D 2.1</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>Scenario 01 – Delays due to changes in the airport configuration</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>Scenario 02 – Delay generation due to staffing problems at one airport</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>Scenario 03 – Restrictions in ATC capacity</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>Scenario 04 – Trajectories modifications due to adverse weather</td>
<td>11</td>
</tr>
<tr>
<td>5</td>
<td>SPECIFICATION OF DATA SOURCES</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>5.1 PRISME DATA</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>5.1.1 PRISME</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>5.1.2 The ALL-FT+ data set</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>5.2 METAR DATA</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>5.3 ATATURK WEATHER DATA</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>5.4 ACC HEADLINES</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>5.5 ENVIRONMENTAL DATA SET</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>5.6 EUROCONTROL ATFM – NOP</td>
<td>20</td>
</tr>
<tr>
<td>6</td>
<td>ANNEXES</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>6.1 DATA CONFIDENTIALITY AGREEMENT ON THE USE OF ALL FT+ DATASET</td>
<td>21</td>
</tr>
</tbody>
</table>
ABSTRACT

This document is deliverable D 2.1 "Description of the integrated database schema", which, as described in the Description of Work (DoW), is part of WP2, with 18 person-months involved in its tasks. The following report delivery date is T0+9 which corresponds to the 1st of February, 2013. It describes the data collection, specifying its functionalities for the Resilience2050 project.
1 INTRODUCTION

Resilience2050.eu is a collaborative project funded through the FP7 AAT Call 5, topic AAT.2012.6.2-4: Identifying new design principles fostering safety, agility and resilience for ATM.

The project aims to:

- Develop adequate mathematical modelling and analysis approaches to support systematic analysis of resilience in ATM scenarios, both in normal operations as well as in disturbances;

- Develop metrics to systematically define resilience, and address related concepts such as "Responding", "Monitoring", "Learning" and "Anticipating". This work will result in a Resilience Analysis Framework (RAF 2050), to enable the definition of new ATM design principles fostering safety, agility and especially resilience.

- Provide an extensive overview of human contributions to resilience in current ATM;

The document structure presents, through a top-down approach, the top level Resilience2050 project information, the role of Work Package 2 within the project and, finally, the work performed under task D2.1 and the description of the different data sets chosen for the project. Concretely:

- Section 2 explains the main aims of WP2 in the overall Resilience2050 project.

- Section 3 consists of the description of D2.1 in relation to the WP2 frame and the general procedure in this work package, in addition to the explanation of the links between D2.1 with the rest of the deliverables in WP2 (D2.2 and D23).

- Section 4 specifically describes the goals and challenges of D2.1, with particular attention to the data requirements which ensures alignment with D2.2 tasks.

- Section 5 is the core of the document, providing a technical description of the different data sets. It includes distinct sub-sections for each of the following: PRISME data, METAR data, Ataturk Weather data, ACC Headlines, DDR Environmental Data Set and Eurocontrol ATFM-NOP.

- Annex I consists of the data confidentiality agreement on the use of the ALL_FT+ dataset.

The project is carried out by an international consortium composed of: The Innaxis Research Institute, (Project Coordinator, Spain), Deutsches Zentrum für Luft- und Raumfahrt e.V (DLR, Germany), Universidad Politécnica de Madrid (UPM, Spain), Nationaal Lucht- en Ruimtevaartlaboratorium (NLR, Netherlands), Istanbul Teknik Üniversitesi (ITU, Turkey), Devlet Hava Meydanlari Isletmesi Genel Müdürlüğü (DHMI, Turkey) and King's College London (KCL, UK).

The project was launched on the 1st of June, 2012 and will last 36 months.
### 1.1 Acronyms and Terminology

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAT</td>
<td>Aeronautics and Air Transport</td>
</tr>
<tr>
<td>ANSP</td>
<td>Air Navigation Service Provider</td>
</tr>
<tr>
<td>ATFCM</td>
<td>Air Traffic Flow Capacity Management</td>
</tr>
<tr>
<td>ATFM</td>
<td>Air Traffic Flow Management</td>
</tr>
<tr>
<td>ATM</td>
<td>Air Traffic Management</td>
</tr>
<tr>
<td>CDM</td>
<td>Collaborative Decision Making (Airports)</td>
</tr>
<tr>
<td>CFMU</td>
<td>Central Flow Management Unit</td>
</tr>
<tr>
<td>CORDIS</td>
<td>Community Research and Development Information Service</td>
</tr>
<tr>
<td>CRCO</td>
<td>Central Route Charges Office</td>
</tr>
<tr>
<td>CSV</td>
<td>Comma Separated Value</td>
</tr>
<tr>
<td>D.X.Y</td>
<td>Deliverable X.Y</td>
</tr>
<tr>
<td>DDR</td>
<td>Demand Data Repository</td>
</tr>
<tr>
<td>DHMI</td>
<td>Devlet Hava Meydanları İşletmesi (The General Directorate of State Airports Authority in Turkey)</td>
</tr>
<tr>
<td>DLR</td>
<td>Deutsches Zentrum für Luft- und Raumfahrt (German Aerospace Center)</td>
</tr>
<tr>
<td>DoW</td>
<td>Description of Work</td>
</tr>
<tr>
<td>EATM</td>
<td>European Air Traffic Management</td>
</tr>
<tr>
<td>EC</td>
<td>European Commission</td>
</tr>
<tr>
<td>FP</td>
<td>Framework Programme</td>
</tr>
<tr>
<td>ICAO</td>
<td>International Civil Aviation Organization</td>
</tr>
<tr>
<td>INX</td>
<td>The Innaxis Foundation and Research Institute</td>
</tr>
<tr>
<td>ITU</td>
<td>Istanbul Teknik Universitesi (Istanbul Technical University)</td>
</tr>
<tr>
<td>KCL</td>
<td>King's College London</td>
</tr>
<tr>
<td>METAR</td>
<td>Meteorological Aerodrome Report</td>
</tr>
<tr>
<td>NLR</td>
<td>Nationaal Lucht en Ruimtevaartlaboratorium (The National Aerospace Laboratory – The Netherlands)</td>
</tr>
<tr>
<td>NOP</td>
<td>Network Operation Plan</td>
</tr>
<tr>
<td>PRISME</td>
<td>Pan-European Repository of Information Supporting the Management of EATM</td>
</tr>
<tr>
<td>QNH</td>
<td>Atmospheric Pressure (Q) at Nautical Height</td>
</tr>
<tr>
<td>R&amp;D</td>
<td>Research and Development</td>
</tr>
<tr>
<td>RAF2050</td>
<td>Resilience Analysis Framework</td>
</tr>
<tr>
<td>SES</td>
<td>Single European Sky</td>
</tr>
<tr>
<td>SESAR</td>
<td>Single European Sky ATM Research Programme</td>
</tr>
<tr>
<td>TAF</td>
<td>Terminal Area Forecast</td>
</tr>
<tr>
<td>UPM</td>
<td>Universidad Politécnica de Madrid (Madrid Technical University)</td>
</tr>
<tr>
<td>WP</td>
<td>Work Package</td>
</tr>
</tbody>
</table>
2 WP2 in Resilience2050

While the key objective of the Resilience2050 proposal is to define, analytically, what the concept of "resilience" means in the context of Air Traffic Management, the WP2 Resilience Analysis Framework (RAF2050) precisely aims to develop a data-driven analytics framework to characterise uncertainty in Air Traffic Management. In this context, WP2 structures other activities from other work packages, as it is key to other activities in the Resilience2050 project; in particular:

- WP3.1 'Creation of the Multi-layered Resilience Representation', will require input from WP2.1 'Description of the integrated database schema'.
- WP3.2 'Resilience Metrics for the European ATM', will require input from WP2.2, the resilience-mining task.
- WP5.1 The definition of the ATM stress tests will require input from the WP2.2 Resilience Mining task.

These relationships with other work packages are shown in the diagram below, representing (partially) the relationships between packages:

![Figure 1 Relationship between work packages](image)

The ultimate goal of this WP is to develop a data-driven framework capable of identifying patterns from the ATM data available. While some research fields have a consistent approach to data collection and data mining, the Air Traffic Management field lacks a unified approach to data availability across geographical boundaries and time scales or integration across different sub-systems (e.g. data comes from different sub-systems representing the same data identity). This lack of integration obviously impacts the data mining techniques available and, as at the time of writing, all research projects that lean on a data-driven concept begin with an analysis of the different data sources and require intense work on clarifying how they can actually support the research concept.

Resilience2050 is not an exception and, therefore, WP2 should identify those data elements that could appropriately represent the ATM Resilience concept. As the "resilience concept" itself is being developed within the project, the research challenges for WP2 are multiple: there is a need to ascertain that the data sources that will support the concept, to understand the availability of the data, its limitations and the constraints of its use (e.g. potential limitation on temporal resolutions) and to develop the data mining concept that will extract the appropriate metrics.
The first task within the WP2 "Resilience Analysis Framework" is D2.1: "Description of the integrated database schema" which focuses directly on data integration and specification and the building of a database architecture.

Technically speaking, D2.1 has set the basis for D2.2 -Description of the ATM patterns and insights discovered- or simply "resilience mining", which will lead afterwards to D2.3 -Specifications and procedures to support the design of a resilience data mining service. The appropriate data search and integration exercise done for D2.1 will be crucial for effective data mining in D2.2 and for viable outcomes in D2.3. Furthermore, D.2.2 results will also be required by other WPs, as described in section 1 of this document.

The tasks of D2.1 include the collection of data and specification of the data management procedures to be used within the research project. Functionalities for processing of raw ATM data and description of the temporal properties of historical data should be developed within this task and will play an important role in the progress achieved during WP2, specifically with D2.2 and D2.3.

Data in aviation is not available freely and with unconstrained permission. On the contrary, it usually requires the attainment of appropriate permissions for its use, delivery and storage under a secured environment. Other tasks aiming to make the data perfectly secure and to maintain the confidentiality of the information are also required.

Therefore, a number of activities looking into securing the availability of the data have to be undertaken. In particular the Resilience2050 team requested that Eurocontrol make available some of the data sources, in particular the ALL_FT+ dataset. Eurocontrol granted permission for the use of this dataset and provided the dataset. The availability of this dataset under the Resilience2050 project is of very high value to the project and it is described in detail in this deliverable. The data is, at the time of writing, stored in secured servers at Innaxis and the University Polytechnic of Madrid. In this context it is worth mentioning the confidentiality agreement to which the project has committed; it is documented in Annex I of this deliverable.

Other data sources that will be available for this project (METAR data, ACC Headlines, Ataturk weather and DDR Environmental data set) follow similar privacy guidelines, which will be applied, not just during D2.1, but also during all of WP2 and the Resilience2050 project.

The specific objectives of D 2.1 and challenges achieved in the preparation of this deliverable with its corresponding tasks, can be found in the following section, Section 4.
4 Objectives, challenges and requirements of D 2.1

While Section 2 details the role of WP2 in the whole project, and Section 3 outlines how this deliverable is linked with the rest of the WP2 deliverables and tasks, the aim of this Section is to provide an accurate and specific description of the target objectives, and the challenges overcome and requirements found during the development of D2.1 and its corresponding tasks.

The first subsection, involving the objectives and challenges of D2.1, describes from an overall point of view the goals and aims of the data sets.

The second subsection, explains the technical requirements, in addition to including a brief explanation of the scenarios studied in this WP2. For this component, extensive meetings with UPM, responsible for D2.2, have been held in order to achieve full alignment between both deliverables.

4.1 Objectives and Challenges

The objectives and challenges of D 2.1 are as follows:

D 2.1 "Data integration and Specification of the Database Arquitecture" is the first technical deliverable remitted in the development of the Resilience2050 project. Although work is being carried out in parallel in WP1 concerning the Resilience concept, D2.1 is the first scientific deliverable covering the technical data that will be used in the project. Hence, it includes the descriptions of the different data collections that, as of today, the consortium consider necessary for fertile project progress, and which are described in detail in the following subsection.

The data collection has been provided with additional information about the data sources, including qualitative and quantitative data properties and a description of the temporal properties of historical data. The material included not only covers granular ATM data such as Ataturk weather data but also the provision of major airport and general aviation performance data like PRISME or METARs. Aviation data has many limitations, as mentioned in previous sections, thus, an in-depth analysis of the data and the technical limitations of its use was necessary.

The tasks corresponding to this deliverable have provided a robust and safe environment for the accommodation of the ATM data: Data must be perfectly secured throughout the project. All the material stored must of course be kept by means that ensure the confidentiality of the information.

Another key challenge has been the integration of the different data sets that will be used, so that, once achieved, all the material could set the basis for future work efforts. Different sources providing the same information have been checked to determine that they are neither redundant nor improper.

All data should be in a suitable digital format, in order to simplify the data mining tasks included in D2.2. Specifically, data sets including natural language texts or images should be avoided.

Last but not least, the effective collaboration of all the consortium partners in setting data requirements and in carrying out the data search itself have been, given their professional origins, of key importance. Research centres, ANSPs and Universities provided different and complementary points of views from their different backgrounds. Hence, particular requirements imposed, taking in to account the future different WPs tasks, have played an essential and productive role in this D2.1

4.2 Requirements of D 2.1

Throughout the process of defining the integrated database schema, special consideration has been devoted to the data needs associated with future tasks: mainly D2.2 and the identification of ATM patterns through historical data mining.
The connection between D2.1 and D2.2 is clear, and requires a complete alignment in order to avoid unnecessary effort now or in the future. For this reason, several meetings were organized, aimed at defining the requirements imposed by D2.2 from the data perspective.

While it was impossible, at this stage, to analyse all requirements associated to the data mining scenarios of D2.2, as many of them have not been defined yet, a general overview is presented here. The overview is based on the four scenarios that have already been defined; yet, a special effort has been made to ensure the generality of such requirements, so that they will cover most of the D2.2 work.

**Scenario 01 – Delays due to changes in the airport configuration**

This scenario will focus on the relation between the appearance of delays at take-off and landing, and the procedures for configuration changes at different airports. Configuration changes are required in order to operate within the mandatory safety envelopes, as only a small (< 10 kts) amount of tail-wind is permitted during take-off and landing, and therefore the active runways should be chosen according to the evolution of the wind through the day. As a configuration change requires the rerouting of all aircraft (both on the ground and in the air), we expect this to increase the delay suffered by flights; furthermore, we expect this delay to vary from airport to airport, as it will depend on runways and organization procedures.

There are three elements required in this scenario:

- Identification of changes in airport configuration; this will require an exact knowledge of the trajectory of the aircraft, both at take-off and at landing.
- Assessment of the additional delay introduced by the configuration change; for this, information about the Actual Off-Block Time (AOBT) of each aircraft will be required.
- Additionally, it would be useful to relate the appearance of configuration changes with variations in the direction of dominant winds; therefore, a record of the weather conditions near the main airports is required.

**Scenario 02 – Delay generation due to staffing problems at one airport**

An airport can have staffing problems due to several causes: from large-scale events such as strikes, to simultaneous absences due to illness. When this happens, the number of aircraft moving on the ground or close to the airport should be reduced, in order to maintain the required safety levels. We expect this to significantly impact the mean delay of departing and arriving aircraft.

The following requirements have been identified in this scenario:

1. Identification of the days, and possibly of the time windows, in which an airport has had some kind of staffing problem. More generally, this would require a data set comprising special events occurring in the European air transport system.
2. Assessment of the inbound and outbound delays. In turns, this will require information about planned and executed flight trajectories.

**Scenario 03 – Restrictions in ATC capacity**

One of the most important perturbations, suffered daily by customers of the European air transport system, is constituted by delays generated by regulated sectors, i.e., sectors for which the declared capacity is not enough to meet the demand. In order to analyse the importance of such a disturbance, we propose the analysis of the delay generated by each sector when such conditions are met. This can be done by extracting the delay at take-off due to regulation for each flight, divided into two categories: flights crossing the analysed sector, and all other flights. These two delay distributions will be compared;
if different in a statistically significant way, we know that the regulation in that sector is generating a stress in the system, beyond the normal delay expected.

This scenario implies the following requirements:

1. Data set reporting the evolution of the declared capacity of European air sectors.
2. A repository of flight trajectories (both planned and regulated) in order to assess additional delays due to regulations.

**Scenario 04 – Trajectories modifications due to adverse weather**

Due to weather phenomena, flights might suffer different types of modifications; from local rerouting (e.g. for avoiding small regions of convective weather), up to complete cancellations due to extreme events (like volcanic eruptions). Furthermore, although flights may not be directly rerouted, adverse weather may cause a drop in the capacity of sectors.

In order to assess the effects of different kinds of adverse weather, the following data are required:

1. Data sets codifying the evolution of weather over the European air space, with enough spatial and temporal resolution to be matched with flight trajectories.
2. A repository of flight trajectories, including planned, regulated and real.
5 Specification of Data Sources

This last section of D2.1 is the core of the deliverable as it includes the description of the integrated database schema. Descriptions of the different data sets that will be involved in the Resilience2050 project, are provided. These are:

- PRISME data
- METAR data
- Ataturk weather data
- ACC headlines
- DDR Environmental data set
- Eurocontrol CFMU - NOP

In each of the data sets a specification of the data is given:

- Data source with an explanation of how the information access has been achieved and some brief information about the original data owner and stakeholders involved.
- Data size: quantity and quality of the data
- Data storage
- Files size
- Data time frame

5.1 PRISME data

The main raw data source in regard to trajectories and air transport performance that will be used in the Resilience2050 project comes from PRISME.

5.1.1 PRISME

The task of PRISME (the Pan-European Repository of Information Supporting the Management of EATM (European Air Traffic Management)), a EUROCONTROL group, is to develop and maintain an integrated ATM data warehouse, i.e., to collect, clean and store data of air operations in the European airspace. This information is then shared with different organisms, such as, for instance, ICAO (International Civil Aviation Authority) or the European Commission, for the preparation of studies and reports.

The data handled by this group come from many different sources, the most important of them being all of the national Air Navigation Service Providers, and EUROCONTROL itself through CRCO (Central Route Charges Office). It is worth noting that, due to this heterogeneity of sources, the information arrives in different formats, to be then processed and homogenized by PRISME. This simplifies the process of achieving a certain standardisation of data in pan-European studies, since getting data from different countries would imply not only different data acquisition procedures, but also data formatting and standardisation tasks.

5.1.2 The ALL-FT+ data set

One of the most relevant sets of data managed by the PRISME group is related to flight trajectories. This data set encodes different types of information for each flight crossing the European airspace, including those intercontinental flights that overfly European airspace. Among others, the most relevant information for the Resilience2050 project included is:

- General information about the flight: identification of the flight and of the aircraft (flight number, aircraft type and registration mark), departure and destination airports, and airline operating the flight.
- Information about regulations (usually, the most penalizing one) assigned to the flight.
• Complete trajectory followed by the aircraft.
• Miscellaneous information, such as, for instance, airport Collaborative Decision Making (CDM) data if applicable.

The data set analysed in this document covers the whole European airspace, from the 1st of March, 2011, up to the 1st of January, 2012. The Resilience2050 project was launched during mid-2012 and this dataset is the one currently available to the project. ATM information has been recorded in this format only since the 1st of March 2011. Later in the Resilience2050 project the team will consider updating to a wider dataset, probably covering at least two full years, from the 1st of March, 2011 to the 1st of March, 2013.

While the overall quality of the data set was excellent, Innaxis identified several drawbacks. Probably the most important was related to the profiles of flights updated with radar information; a high percentage of such trajectories were not updated with actual radar information, and the “radar-based profile” was just the planned trajectory. This is due to the way the ALL-FT+ data set is constructed: it is built by merging information provided by different ASNPs across Europe, with heterogeneous qualities (temporal and spatial resolutions) and coverage. Furthermore, even when flight trajectories are updated with radar information, the mean temporal resolution is restricted to a point every 90 seconds. In the context of the Resilience2050 project, it will be necessary to check whether this source of data fulfils the requirements of precision and reliability.

In the following figures, some of the key characteristics of this data set are graphically represented. Specifically, the evolution of the global number of flights included in the data is plotted in Figure 2. In this figure the reader can appreciate the strong variability, corresponding to weekly and annual seasonality.
Figure 2 Evolution of the global number of flights included in the data

Furthermore, the reader can see the density of flights over European airspace in Figure 3, where a map of this is plotted. Although results are consistent with what is normally expected overall, it should also be noted that there are some regions of the airspace displaying abnormal characteristics. For instance, there are regions for which there seems to be a lower traffic density than expected (e.g., over France); this effect is due to the plotting procedures (which consider only the beginning of each segment in the calculation of the density of flights in a region), and to a lower temporal resolution of the radar profiles. Furthermore, it can be seen that some flights disappear in the Atlantic Ocean, west of France; this is a known issue of compatibility of the equipment providing the radar tracks in that region.

Figure 3 Density of flights over European airspace

Figure 4 is also provided in order to justify these previous explanations. It comes from ICAO and can be a great source of comparison despite its obvious differences with the ALLFT+ image:

- ICAO flights are represented as origin-destination lines: real trajectories are not the ones given in comparison with ALLFT+ image.
- ICAO represents 2010 traffic flow meanwhile ALLFT+ considers 2011 operations.
- There is a different colour code density of flights.
5.2 METAR data

One of the sources of information most important for pilots, especially for approach and departure phases, are METARs (METeorological Aerodrome Report). These reports are prepared by all major airports every 30 minutes, and include all facts relevant to low altitude air navigation and landing/take-off operations. Furthermore, this information strongly affects real operation, and is therefore expected to be related with the appearance of safety events.

As part of the data management activities of the Resilience2050 project, we have collected the METARs corresponding to the 40 busiest airports in Europe, ranked by total passengers per year in 2010. The number of available METAR reports for the first 10 of them is reported in Table 1.
Table 1 Information available in the METAR datasets

<table>
<thead>
<tr>
<th>Airport name</th>
<th>Country</th>
<th>ICAO code</th>
<th>Number of METARs available</th>
</tr>
</thead>
<tbody>
<tr>
<td>London Heathrow Airport</td>
<td>United Kingdom</td>
<td>EGLL</td>
<td>20600</td>
</tr>
<tr>
<td>Paris-Charles de Gaulle Airport</td>
<td>France</td>
<td>LFPG</td>
<td>17849</td>
</tr>
<tr>
<td>Frankfurt Airport</td>
<td>Germany</td>
<td>EDDF</td>
<td>24952</td>
</tr>
<tr>
<td>Barajas Airport</td>
<td>Spain</td>
<td>LEMD</td>
<td>19535</td>
</tr>
<tr>
<td>Amsterdam Airport Schiphol</td>
<td>Netherlands</td>
<td>EHAM</td>
<td>20748</td>
</tr>
<tr>
<td>Leonardo da Vinci-Fiumicino Airport</td>
<td>Italy</td>
<td>LIRF</td>
<td>20585</td>
</tr>
<tr>
<td>Munich Airport</td>
<td>Germany</td>
<td>EDDM</td>
<td>17811</td>
</tr>
<tr>
<td>Atatürk International Airport</td>
<td>Turkey</td>
<td>LTBA</td>
<td>18892</td>
</tr>
<tr>
<td>Gatwick Airport</td>
<td>United Kingdom</td>
<td>EGKK</td>
<td>17733</td>
</tr>
<tr>
<td>Barcelona El Prat Airport</td>
<td>Spain</td>
<td>LEBL</td>
<td>18937</td>
</tr>
</tbody>
</table>

Note that, even for all these airports, a METAR is prepared each 30 minutes. The number of METAR reports available for each airport may vary; this is due, among other reasons, to the availability of such information and the existence of special METARs prepared to inform stakeholders about adverse events.

For each one of these reports, the available information is reported in Table 2.

Table 2 Information available in the METAR reports

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Format</th>
</tr>
</thead>
<tbody>
<tr>
<td>Date and time</td>
<td>Date and time corresponding to the observation</td>
<td></td>
</tr>
<tr>
<td>Temperature</td>
<td>Temperature, in ºC</td>
<td></td>
</tr>
<tr>
<td>Humidity</td>
<td>Relative humidity measured at the airport.</td>
<td>Integer number between 0 and 100.</td>
</tr>
<tr>
<td>Pressure</td>
<td>QNH (barometric pressure adjusted to sea level) at the airport.</td>
<td>Integer number, usually quoted between 950 and 1050.</td>
</tr>
</tbody>
</table>
The information contained in the METARs for the European airspace has been collected through a custom Web Spider software, created specifically for this task, whose task was to automatically download the available information from different web sites. Once this information was retrieved, the program further converted the format of each source (in order to obtain a coherent representation of the information) and saved the result in different files, one for each airport. It should be noted that, due to the non-official sources used to retrieve the data, this information contains several errors, mainly missing METARs.

5.3 Ataturk Weather Data

As the Resilience2050 consortium opportunely includes DHMI (The General Directorate of State Airports Authority in Turkey), there has been easy access to the complete weather data set of Ataturk airport. This source of information is in any case complementary to the general PRISME and METAR data sets previously explained in this section, which included meteorological and trajectory data from all over Europe, where this Turkish region was also contained. Nevertheless, the ultra high level of detail of the weather data, especially regarding "bad weather conditions", will be of great value in the development of the Resilience2050 project.

DHMI, which represents the Turkish State Airports Authority, is associated with the Ministry of Maritime Transport and Communications. DHMI’s Air Navigation Department, mostly with the participation of its ATM R&D Management, are the participants who have mainly provided Resilience2050 with all the detailed information in regard to Ataturk airport, through the Meteorological Department (METEOROLOJI GENEL MÜDÜRLÜĞÜ)

The information is compiled in 16 different Excel (xls) files:

- Daily_maximum_wind_speed_(10_m.de)(m+A?sec)_and_direction(yon_adı) [38Kb]
- Daily_Snow_thickness_(cm)_ [27 KB]
- Daily_Snowy_days [16KB]
- metar_data_2009_2012 [1.8MB]
- Monthly_closed_days_number [30KB]
• Monthly_Foggy_days [30KB]
• Monthly_maksimum_Snow_thickness_(cm) [30KB]
• Monthly_maksimum_Wind_Speed_(10_m.de)(m_sec)_and_direction(yon_adi) [31KB]
• Monthly_maximum_wind_speed_(m_sec) [30KB]
• Monthly_Snow_covered_days [30KB]
• Monthly_Stormy_day_number [30KB]
• taf_data_2009_2012 [244KB]
• Temperarature_Pressure_data_2009_2012daily [74KB]
• Temperature_pressure_data_2009_2012 [1.7MB]
• wind_data_2009_2012 [2.5MB]
• wind_data_2009_2012daily [38KB]

There is both daily and monthly information from the Ataturk airport. Although most of the data covers 2009 to 2012, there are some files where the material included covers that time frame only partially. In the case that there is any additional data set needed for this Airport during the project development, it could be provided by DHMI. The following data metrics are contained in the files:

1. Temperature data (temp, temp Dew Point, average, minimum and maximum included)
2. Pressure data (pressure, pressureQNH, pressureQNHinch)
3. Wind data, including maximum wind speed (m/s) and direction (3 letter format: eg: NNE)
4. Stormy days data
5. Foggy days data
6. Snowy days data, including snow thickness in those snowy days (cm)
7. TAF data (Terminal Area Forecast)

5.4 ACC Headlines

With the term ACC Headlines we refer to a dataset containing information about special events affecting the European airspace. This dataset has been provided by Eurocontrol, and the events therein contained are those that have been published in the CFMU reporting (now called DNM reporting) from the 1st July, 2010 to the 31st of December, 2011.

The original format of the dataset is a comma separated value (CSV) file, where each row corresponds to a different day, and all the events corresponding to that day are reported in a textual form. For instance, the row corresponding to the 3rd of July, 2010 starts with:

03-Jul-10,"LTAI radar problems. LPPR low visibility. LFPG CB activity. LIML closed due to oil on the RWY. ...

Clearly, such a format is not suitable for a data-mining process, as any computation will require input information to be in a raw format, therefore, the first step has been the development of an algorithm that simplifies the codification of the information.

The final representation of the ACC Headlines has been organized in records, each one representing a single event, and that include the following information:

• Date of the event;
• Airport affected by the event.

5.5 Environmental Data Set

The Demand Data Repository (DDR) is a tool, developed by EUROCONTROL inside the DMEAN programme, which gives users access to historical and forecasted air traffic demand, with the main aim of supporting more efficient operation planning. It also contains relevant information about the status and structure of the European airspace, that is, Environmental information.

The available sample of DDR data spans from the 1st of January to the 31st of December, 2011. Information is not codified in the same way during the whole period; more specifically, data change according to the 28 day AIRAC cycle, and include changes in the actual configuration of the airspace, and especially in the number of sectors.

Among the information composing this data set, the following are of interest for the Resilience2050 project:

1. **Sector capacity**, containing information about the declared capacity of each sector in the European airspace. Information is organized in events, where each event defines the capacity of a single sector in a given time window (notice that different capacities may be defined for different moments of the day, thus drastically increasing the number of events available in this dataset. Each event, therefore, contains four fields of data:

   • Name of the sector;
   • Initial and final time;
   • Declared capacity (a value of 999 represents no limitation in capacity).

2. **Sector configuration**, reporting the declared configuration of each one of the sectors defined within the European airspace. Its structure is reported below:

   • Name of the zone in which the sector is included (normally, the name of the corresponding ACC);
   • Name of the configuration;
   • Name of the sector.

   It is worth noting that little knowledge of this dataset is included as we have no information about what each configuration implies for the real operation of aircraft. In other words, there is no way of understanding the differences between “CONF1” and “CONF2”. In any case, this information will be included in the data-mining process, and the relation between the changes of the configuration and the appearance of safety patterns will be analysed.

3. **Active runway**, i.e. information about the active runways of each airport, including the initial and final time of the period of activation. The information that is extracted is described below:

   • ICAO code of the airport;
   • Active runway (identified as a pair of numbers);
   • Initial and final time of the period of activation. The other fields included in the dataset have been discarded, as no documentation has been obtained, and their meaning has not been completely clarified.
5.6 Eurocontrol ATFM – NOP

Despite the fact that this data set has not been used by itself, Eurocontrol ATFM - NOP is included in this deliverable as a data source.

Eurocontrol NOP (Network Operations Portal) is a free access webpage where there are numerous data elements regarding the ATFCM Network situation in the European airspace. The information that will be used for the Resilience2050 project will be minimum, as only the list of disturbances will be used - as a source of classifying the disturbance list, which will be studied in depth in the next deliverables.

The complete (and raw) disturbance list is as follows:

- Weather
- ATC Staffing
- Aerodrome Capacity
- ATC Capacity
- Airspace Management
- Special event
- ATC Routeings
- ATC Equipment
- ATC Ind Action
- Accident/Incident
- De-Icing
- Environmental issues
- Equipment non-ATC
- Ind Action non-ATC
- Others
6  Annexes

6.1  Data Confidentiality Agreement on the use of ALL FT+ dataset


Due to the inherent confidentiality restrictions of the different datasets, the Resilience2050 Consortium follows the following data management policies for all datasets being used. Each entity forming the consortium may sign different data agreements depending on the owner of the data. These policies ensure the maintenance of data confidentiality and allows data owners to understand the framework in which the data is to be used, making it easier for them to authorize within their respective organisations the use of the data for research goals. The following policies make up the actual Data Management Agreements signed between the parties involved:

Regarding the PERIOD OF PERFORMANCE

This agreement shall be effective on the date indicated on the last page, when signed by both parties, and shall be valid until the expiration or termination of the contract identified in the Purpose.

Nevertheless, the obligations of the Receiver will apply up to ten years after termination of the present agreement.

Regarding the DESCRIPTION OF DATA

Data in the context of the present Agreement are those data sets provided by the TBD for the purpose described in Section II.

Data will be organised in different datasets by the Provider.

Each Dataset provided to the Receiver will consist of a certain number of files.

The content and the format of the files will be those necessary for the purpose described in Section II.

The Provider has no liability for the content of the Data, nor for the damages that may result from the use of the Data.

Regarding the ACCESS TO DATA

A. METHOD OF ACCESS AND TRANSFER

Unless otherwise stated, the datasets will be provided and stored on portable discs by the Provider.

The Receiver will formally designate to the Provider a maximum of two representatives allowed to receive the portable discs. Receiver will be responsible for the physical, secure collection of portable discs at the Provider’s office.

Portable discs will not be returned by the Receiver and will not be reused for any other purpose. The Receiver will securely store the portable discs up to the end of contract or securely destroy them.

Regarding the PERSONS HAVING ACCESS TO DATASETS

Only members of the Receiver organization’s staff who specifically require access to such datasets in the performance of their assigned duties shall have access to these datasets.
Prior to any dataset transfer under this Agreement, all staff members who have access to the data will be notified of the present Agreement and shall sign a declaration certifying the understanding of the conditions contained in this Agreement. These declarations will be kept by the Receiver.

**Regarding the SECURITY OF DATA IN THE RECEIVER ENVIRONMENT**

All reasonable precautions, including administrative, technical, personnel, and physical measures, shall be taken to secure the datasets against loss, misuse, unauthorized access, disclosure, alteration, uncontrolled destruction, and theft.

The Receiver will keep track of dataset uses and of staff members authorized to access the datasets and will report this information back to the Provider when requested, including the signed declarations.

At termination of the contract indicated in Section II, the Receiver will securely erase all data received and its copies, and will send a document to the Provider certifying that all data have been securely erased.

**Regarding the CONFIDENTIALITY**

A. REGULATIONS COVERING CONFIDENTIALITY OF DATA

The use and disclosure of information obtained under this Agreement shall be subject to National Regulations and to the specific obligations of the contract indicated in Section II.

The Receiver shall maintain the confidentiality of any information that may, in any manner, violate the commercial secrecy of any particular company or organization.

The Receiver shall not link provided data with any other individually identifiable data from any other source, unless specifically authorized in the contract identified in Section II.

B. NON-DISCLOSURE OF DATA

The Receiver staff shall not disclose, in whole or in part, datasets described in this agreement nor the possession of those datasets to any individual, company or agency not specifically authorized in the contract identified in Section II.

**Regarding the PAYMENT**

No compensation will be required by either party.

**Regarding the COPYRIGHT**

Registration of copyright is not allowed, neither of the datasets provided nor of any results achieved thanks to its access.

Commercial use of data included in the datasets and commercial use of information derived from the data included in the datasets is explicitly forbidden.

**Regarding the WAIVER OF DEFAULT**

Waiver of any default shall not be deemed to be a waiver of any subsequent default.

Waiver of a breach of any provision of the Agreement shall not be deemed to be a waiver of any other or subsequent breach and shall not be construed to be a modification of the terms of the Agreement unless stated to be such in writing, signed by the Contracting Officer, or Executive Director of (Entity) and attached to the original Agreement.
Regarding the ALL WRITINGS CONTAINED HEREIN

This Agreement contains all the terms and conditions agreed upon by the parties.

No other understandings, oral or otherwise, regarding the subject matter of this Agreement shall be deemed to exist or to bind any of the parties hereto.

--- End of Document ----