D2.1 Data management and sources

Deliverable 2.1

Domino

Grant:	783206
Call:	H2020-SESAR-2016-2
Topic:	SESAR-ER3-06-2016 ATM Operations, Architecture, Performance and Validation
Consortium coordinator:	University of Westminster
Edition date:	14 June 2018
Edition:	01.00.00

SESAR JOINT UNDERTAKING

Authoring & Approval

Authors of the document

Name/Beneficiary	Position/Title	Date
Tatjana Bolić / University of Trieste	Project member	12 June 2018
Lorenzo Castelli / University of Trieste	Project member	12 June 2018
Luis Delgado / University of Westminster	Project member	12 June 2018
Gérald Gurtner / University of Westminster	Project member	12 June 2018
Graham Tanner / University of Westminster	Project member	12 June 2018

Reviewers internal to the project

Name/Beneficiary	Position/Title	Date
Lorenzo Castelli / University of Trieste	Project member	13 June 2018
Graham Tanner / University of Westminster	Project member	13 June 2018

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

Name/Beneficiary	Position/Title	Date
Andrew Cook / University of Westminster	Project coordinator	14 June 2018

Rejected By - Representatives of beneficiaries involved in the project

Name/Beneficiary	Position/Title	Date
N/A		

Document History

Edition	Date	Status	Author	Justification
01.00.00	14 June 2018	Release	Domino Consortium	New document for review by the SJU

Domino

NOVEL TOOLS TO EVALUATE ATM SYSTEMS COUPLING UNDER FUTURE DEPLOYMENT SCENARIOS

This deliverable is part of a project that has received funding from the SESAR Joint Undertaking under grant agreement No 783206 under European Union's Horizon 2020 research and innovation programme.

Abstract

This deliverable presents the approach of Domino for the data management, and details the data sources considered.

Table of Contents

List of tables

Executive summary

Domino will analyse the coupling of the elements of the air traffic management (ATM) system, and how the relationship between the elements and their criticality change when different mechanisms are implemented, from both the flight and passenger perspective. Having sufficient data to be able to model the different interactions is paramount. Domino will develop an agent-based model (ABM) to represent the interaction between the elements in the system and the behaviour of the different actors in the system. As the goal is to assess the ATM system from both the flight and passenger point of view, the data regarding the system infrastructure, traffic (flights) and passenger itineraries are required to be stored, and prepared for the use in the Domino model. Finally, Domino will use complexity science tools to analyse the results of the model, thus the management of the outcome of the modelling is equally important as the management of the input data.

The data Domino will acquire, elaborate and manage is based on the data requirements, coming from the modelling needs. The data that Domino will manage can be categorised as follows:

1. Traffic and delay data (e.g. trajectories, causes and amounts of delay);
2. Passenger data (e.g. passenger itineraries);
3. Airspace environment data (e.g. airport and airspace capacity);
4. Other data (e.g. flight cost data, route charges and airline alliances).

Most of the datasets have already been acquired, and just need to be loaded into the database. The database is already set-up, hosted by the University of Westminster. Secure access to the database by the Domino partners is also in place and all the partners working on the project have access to it.

Next steps involve the population of the database with the (new) data relevant to the Domino project, i.e. data from September 2017. As the project progresses, new tables will be added to store intermediate data input and the modelling results, applying a versioning scheme, in order to keep track of the project progress.

1 Introduction

Domino will analyse the coupling of the elements of the air traffic management (ATM) system, and how the relationship between the elements and their criticality change when different mechanisms are implemented, from both the flight and passenger perspective. Having sufficient data to be able to model the different interactions is paramount. Domino will develop an agent-based model (ABM) to represent the interaction between the elements in the system and the behaviour of the different actors in the system. As the goal is to assess the ATM system from both the flight and passenger point of view, the data regarding the system infrastructure, traffic (flights) and passenger itineraries are required to be stored, and prepared for the use in the Domino model. Finally, Domino will use complexity science tools to analyse the results of the model, thus the management of the outcome of the modelling is equally important as the management of the input data.

Moreover, the Domino consortium is formed by five institutions (partners) located in four different countries. Different partners are responsible for the development of different parts of the model and the analysis of the outcome data. This presents a particular set of requirements in terms of data accessibility, security and reliability that will be detailed in the following sections.

The acquisition and management of data used in Domino is performed in the WP2 of the project.

2 Data requirements

Domino is interested mainly in the relationships of different elements of the systems, in particular flights/company, aircraft, and passengers. The flight/passenger view is of particular importance. The consortium agreed on the list of data requirements needed for the modelling and the results analysis. Most of the data have already been collected, even if further cleaning is required, as explained in Section 3.

Domino has three mechanisms that will be modelled to create the different case studies, as explained in Deliverable D3.1. For each of the mechanisms, and scenarios involving them (see D3.2), Domino will carefully assess the necessary ingredients (systems, subsystems) required to capture the main network effects (see D3.1 for a first list and D4.2 for further details). The minimum set of actors and sub-systems is defined (see D3.1 and D4.1), as well as the details of the case study mechanisms (see D3.1), thus the initial data requirements are already known and are described in the following four subsections.

2.1 Delay management strategies

Delay management strategies combine all the different activities that airspace users (AUs) can do in response to delay. The following strategies are to be taken in consideration:

1. Trajectory and speed adjustments (e.g., flight level capping, re-routing or dynamic cost indexing). In order to model and assess these actions, trajectory and aircraft performance data are needed. Computing the fuel consumption in different cases will be done using the BADA model (EUROCONTROL, 2018b).
2. Consideration of passengers' connections and the options available to airlines to minimise their impact on their cost (e.g., actively waiting for passengers at a hub or recovering delay of an inbound flight to ensure the connections are met). For this, passenger itinerary data are needed. Using information on ticket fares can also help the decision by computing the cost of re-accommodation for the airline.
3. The delay recovery due to small buffers between the arrival of an aircraft and its subsequent departure. In order to include these, the aircraft rotations and scheduled departure times are needed. The rotations will be sourced from the DDR2 data (EUROCONTROL, 2018a), and the scheduled departure times will be extracted from first filed flight plans, so called m0 files (from the Integrated Initial Flight Plan Processing System (IFPS) data when available).

2.2 Flight prioritisation

Flight prioritisation processes can be used by airlines pre-tactically to indicate which flights have a higher criticality (i.e. importance) for them when dealing with capacity-demand imbalances. These techniques are related with User Driven Prioritisation Processes (UDPP). As above, the information on scheduled departure times is important for considering airport slot swapping. Further information needed regards the causes of delay – for example, delay due to capacity restrictions (regulations), or due to airline internal matters, or due to airport issues. In order to calibrate the ABM model, the data on causes of delays will be needed. The consortium has access to CODA data (EUROCONTROL, 2018c) through EUROCONTROL's OneSky portal, and these data contain the basic information on the delay causes. Furthermore, the Air Traffic Flow and Capacity Management (ATFCM) regulation data will be used to create the delay distributions due to capacity restrictions. The airport and airspace capacity data will also be needed, and can be obtained from DDR2.

2.3 Flight arrival coordination

Even if Air Traffic Flow Management (ATFM) slots have been assigned pre-tactically, the sequencing and merging of flows at the airport still needs to be done to account for tactical uncertainty and to maximise runway throughput. Flight arrival coordination will deal with different possibilities to perform these activities building on the E-AMAN concept of operations. For this, detailed trajectory information, departure and arrival times are needed. In particular, uncertainty on departures and arrivals will need to be carefully calibrated. DDR2 traffic data will be used for this, in particular using the last filed flight plan and the actual trajectory. The airport and airspace capacity data will be needed here as well.

Aircraft performance (from BADA) might also be needed when optimising the arrival flows. Similarly, for delay management strategies, passenger itineraries and flight rotations might need to be taken into consideration during the arrival coordination.

2.4 Other

Other data are also required to build the different models. In particular, information on weather (winds) could be useful for the different cost functions of the agents, especially for flights. However, detailed meteorological data will not be required, since estimates of the winds can be extracted from DDR2 data at the adequate level of detail for Domino.

Putting a price on the various delays incurred by the flights is also needed in order to properly derive the behaviours of the agents via their cost function. This will be done using (Cook & Tanner, 2015) updated for 2017[1] (the reference's baseline was 2014). This document and its predecessor, created by UoW, has become a standard in the industry, for instance for the Performance Review Body (PRB) to estimate the total cost of delays in Europe.

[1] Due to time and resourcing constraints, inflationary changes between 2014 and 2017 may be applied to each of the cost of delay components rather than a more in-depth recalculation of the reference values, except fuel costs, which will be explicit.

Various other data might be required to properly calibrate the model. Among them, the Central Route Charges Office (CRCO) charges for the day of operation for every Air Navigation Service Provider (ANSP) will likely be required to properly assess the operational costs. Details on the different alliances and partnerships between airlines will also probably be used in the model, in particular for delay management strategies and for passenger connections. These data are either easy to obtain or are already acquired and consolidated, so the team does not foresee any major acquisition of data apart from those listed above. WP2 runs for most of the project to monitor the needs of the project, acquire more data if needed and manage the model output data.

3 Data sources, acquisition and elaboration

Based on the data requirements detailed in the previous section, here we describe the data sources, acquisition strategy and elaboration of data in order to prepare the input data for ABM modelling. The following data categories have been identified:

1. Traffic and delay;
2. Airspace environment;
3. Passengers;
4. Other.

The sources, acquisition status and elaboration needs are further detailed for each of the categories in the following text. In order to elaborate the data, we will first load the data in their raw form into the database. The database structure is described in Section 4.

3.1 Traffic and delay

Different sources will be used and consulted in order to prepare the traffic data for Domino's models:

- EUROCONTROL's DDR2: flight trajectories from the day chosen as a test day (see the description of the test day selection below) will be sourced from DDR2 data. The trajectories of interest are the last filed flight plans (m1), regulated flight plans (m2) and executed flight plans (m3). Besides the trajectories on the selected day of operations, Domino will source data covering different AIRACs (Aeronautical Information Regulation and Control) to perform statistical analyses when required (for example for the distribution of trajectories). The aircraft rotations for the selected test day will also be acquired from the DDR2 data.
- IFPS initial flight plans: Domino also has access to IFPS data (m0) for the month considered in the scope of the study. These data contain the initial flight plans submitted by airlines, providing information such as the initial estimated off-block time (a good match with the scheduled departure time) and initial desired route (before reacting to congestion/uncertainty).
- Daily ATFCM summary data: contain more detailed information on the regulations that were applied on the day of study. These are obtained from Network Manager (NM) ATFCM statistics.
- CODA summary delay data: will be used to analyse delay and enable realistic delay generation (for all the delays not caused by the ATFM actions).

- CODA taxi times: standard taxi times, published by CODA, are useful to model the time between gate and runway and from the runway to the stand as these values are not present in the DDR2 data, but are required to accurately estimate the arrival and departure delay, and passenger connections.
- BADA performance models: finally, in order to model fuel consumption, Domino will use BADA 4.2 performances from EUROCONTROL. The consortium members have access to BADA 4.2.

3.2 Airspace environment

- EUROCONTROL's DDR2: the DDR2 repository also contains information regarding the airspace environment in terms of sector shapes, sector activations, sector and airport capacities, and basic ATFM regulations. Domino will need the capacity and regulation information, for the level of aggregation that will be determined during the ABM model design.

3.3 Passengers

- Previous itineraries 2010 and 2014: passenger itineraries will be based on previous datasets developed by the University of Westminster for 2010 and 2014. These itineraries will form the basis for the new 2017 itineraries that will be generated for Domino by the consortium.
- GDS data (subject to availability): additional data sourced from a Global Distribution System (GDS) may also be used as inputs to the passenger itinerary generator or for calibration purposes, if the need arises.
- Airline load factors: passenger load factors reported by airlines.
- Airport connectivity information: airport reports on passenger connectivity from ACI EUROPE and individual airports.

3.4 Other

- Cost of delay: cost of delay models developed in-house by the University of Westminster (Cook & Tanner, 2015) will be revised for 2017. As noted in Section 2.4, due to constraints this revision may be based on inflationary changes rather than a more in-depth recalculation of the reference values, except fuel costs, which will be explicit.
- CRCO unit rates: unit rates in effect on the selected test day for every Member State will be required to properly assess the operational costs of modelled flights.
- Airline alliances: details on the different alliances and partnerships between airlines will also probably be used in the models. These data are either easy to obtain or are already acquired and consolidated.

3.5 Summary of data sources

Table 1. Summary of data sources

Category	Datasets	Acquisition status	Elaboration status
Traffic and delay	• DDR2 • IFPS initial flight plans • ATFCM summary data • CODA summary delay data • CODA taxi times • BADA performance models	• DDR2 traffic data available to consortium for different AIRACs • IFPS data available for September 2017 • ATFCM summary data available for 2017 • CODA summary delay data yet to be acquired for September 2017 • CODA taxi times available for the summer season 2017 • BADA - Domino partners in possession of BADA licenses	Need to load the new data into the database (database structure is described in the next section).
Passengers	• Previous itineraries 2010 and 2014 • GDS data (subject to availability) • Airline load factors • Airport connectivity information	• Available in-house (University of Westminster) • Subject to the need and availability • Passenger load factors are yet to be acquired for September 2017 • Connectivity information yet to be acquired for 2017	Need to load the data into the database.
Airspace environment	• DDR2 files containing airport and airspace capacity	• DDR2 airspace environment data available to consortium for different AIRACs	Need to load the new data into the database.
Other	• Cost of delay • CRCO unit rates • Airline alliances	• Available in-house (University of Westminster) • Unit rates available for 2016, 2017, 2018 • Available in-house (University of Westminster), though yet to be updated with 2017 alliance changes	Cost of delay needs an update (mostly on the cost of fuel). Unit rate data needs to be uploaded into database

3.6 Selection of the test day

We have targeted a busy, but not unduly disrupted day in September 2017. Note that Fridays have been selected as they are usually the busiest day of the week. In addition to a French Air Traffic Control (ATC) strike, many days were disrupted by various airline strikes/problems. Ryanair was particularly affected, with nearly a thousand cancellations over the month. There are likely to be other disruptions we are not aware of.

In summary:

- 1st choice: Friday 01 September 2017;
  - Ranked #3 in September 2017;
  - Ranked #5 in 2017;
  - Total ATFM delay quite high (but lower than 2nd choice);
- 2nd choice: Friday 08 September 2017;
  - Ranked #1 in September 2017;
  - Ranked #2 in 2017;
  - Total ATFM delay quite high.

Table 2. September 2017 traffic summary

Rank (month)	Rank (year)	Date	Total flights	Total ATFM delay minutes	Total ATFM strike delay minutes	Total non-ATFM strike delay mins	Total ATFM weather delay mins	Comments
1	2	Fri 08SEP17	37 073	80 611	0	0	29 406	Thomas Cook strike (minor) **2nd choice for test day**
2	3	Thu 07SEP17	36 881	51 601	0	0	7 865	
3	5	Fri 01SEP17	36 798	66 991	0	0	16 419	**1st choice for test day**
4	6	Fri 15SEP17	36 792	90 136	0	0	25 907	Ryanair disruption (number of cancellations unknown)
5	13	Thu 14SEP17	36 313	88 998	0	333	26 492	Ryanair disruption (number of cancellations unknown)
6	16	Mon 04SEP17	36 209	52 571	0	0	12 029	
7	17	Fri 22SEP17	36 193	67 645	0	0	23 541	Ryanair disruption (50 flights cancelled)
8	20	Fri 29SEP17	36 068	78 190	0	0	18 955	Ryanair disruption (56 flights cancelled)
9	27	Wed 06SEP17	35 872	38 003	0	0	7 836	
10	33	Mon 11SEP17	35 681	102 209	19 195	0	32 383	French ATC strike
11	37	Thu 21SEP17	35 595	72 593	15 902	0	6 568	French ATC strike; Ryanair disruption (82 flights cancelled)
12	40	Thu 28SEP17	35 531	46 870	0	0	2 264	Ryanair disruption (48 flights cancelled)
13	42	Tue 05SEP17	35 518	20 146	0	0	144	
14	43	Mon 18SEP17	35 456	70 196	0	0	25 895	Ryanair disruption (65 flights cancelled)
15	50	Wed 13SEP17	35 246	99 841	140	0	52 712	French ATC strike; Ryanair disruption (number of cancellations unknown)
16	55	Wed 20SEP17	35 078	40 942	0	0	2 672	Ryanair disruption (64 flights cancelled)
17	56	Mon 25SEP17	35 064	46 254	0	0	14 460	Ryanair disruption (50 flights cancelled)

Rank (month)	Rank (year)	Date	Total flights	Total ATFM delay minutes	Total ATFM strike delay minutes	Total non-ATFM strike delay mins	Total ATFM weather delay mins	Comments
18	68	Tue 19SEP17	34 712	57 265	0	0	27 895	Ryanair disruption (62 flights cancelled)
19	72	Tue 12SEP17	34 620	188 051	92 803	196	13 123	French ATC strike; Air Berlin disruption (70 flights cancelled); Ryanair disruption (219 flights cancelled)
20	73	Wed 27SEP17	34 594	57 425	0	0	27 090	Ryanair disruption (38 flights cancelled)
21	81	Tue 26SEP17	34 242	44 897	0	0	15 618	Ryanair disruption (44 flights cancelled)
22	83	Sun 03SEP17	34 211	67 864	0	0	7 783	
23	93	Sun 10SEP17	33 814	103 490	0	0	32 291	
24	105	Sun 17SEP17	33 377	70 217	0	0	25 687	Ryanair disruption (80 flights cancelled)
25	114	Sun 24SEP17	33 143	71 900	0	0	13 781	Ryanair disruption (50 flights cancelled)
26	145	Sat 02SEP17	32 092	80 541	0	0	19 964	
27	156	Sat 09SEP17	31 701	108 664	0	0	35 151	
28	164	Sat 16SEP17	31 163	68 026	0	0	13 673	Ryanair disruption (80 flights cancelled)
29	171	Sat 23SEP17	30 896	53 660	0	0	3 965	Thomas Cook strike; Ryanair disruption (50 flights cancelled)
30	186	Sat 30SEP17	29 966	64 335	0	0	10 973	Ryanair disruption (52 flights cancelled)

Sources: NM ATFCM statistics (delays); DDR2 (number of flights)

4 Database infrastructure

All data used in the Domino project are centralised in a single, secure database hosted at the University of Westminster.

4.1 Database access

Due to the various Non-Disclosure Agreements (NDAs) signed by UoW and the other partners for the data, as well as the need to protect the results of the model, access to the database needs to be properly secured. UoW has set-up a database on a virtual machine inside the University cluster, with access password-protected and encrypted with an SSL certificate.

Once logged-in, partners have permission to use the database resources for testing and production. The UoW cluster gives access to easy parallelisation, using more than twenty Central Processing Unit (CPU) nodes and a few Graphical Processing Unit (GPU) nodes too. They are well designed for the kind of mid-range computations that we will require during the production phase of Domino.

Access to data by the different partners is limited considering the different data requirements by the different institutions and subject to having the adequate licencing agreements. The control of data access ensures that possible data corruption is minimised. For instance, UNITS has full writing and reading access to the data, since they are managing the content of the database in Domino while other partners involved in the modelling have read-only access, or can create new tables but not erase any.

4.2 Database structure

The database itself is a MySQL database. MySQL is an open source standard for relational databases all around the world. It is well documented, reliable, and well suited for mid-range databases.

Domino uses the database for two purposes:

- To have standard input data with easy access.
- To store the results of the model(s) in an efficient way.

The structure of the database should be compatible with the following requirements of the models:

- Reproducibility: getting the same output from the same input with the same code.
- Reliability: making sure that the input data has not changed between two runs of the model.
- Consistency: making sure that the input in particular is self-consistent.
- Traceability: making sure that the output data can be linked unambiguously to a given input dataset.

Building-up on data management experience from past projects, Domino will thus use three different types of tables/schemas:

- Some schemas for the primary data, which should never be modified. This includes the DDR2 data for instance and other sourced data (see previous section).
- Some tables/schemas for the secondary data, which are built 'off-line' by some pre-processing codes of the models. These data change with the maturity of the models, and should be versioned.
- Some tables/schemas for the output data, which are the results of the models. Once again, these data change during the project, and should be versioned.

By versioning the secondary data and output data, the project ensures the traceability of the results. While the primary data are in their schemas in the database, all the direct input and output of the model will be centralised in the same schema, called *dominoenvironment* in the database. Note that this schema is used as a placeholder for quite unstructured data, without enforcing the inner consistency of the data via formal relationships. The consistency of the input data is ensured upstream by the pre-processing tools, which should run different tests to this aim.

5 Next steps and look ahead

In terms of data, other than minor 2017 updates to existing data, most of the datasets have already been acquired by the consortium. The only exception is the CODA data on the causes of delays, which are not critical and will be acquired soon via ECTL.

Regarding the preparation of data, the team still needs to load the relevant AIRAC (2017) into the database. This will be straightforward for DDR2 data, and only minor changes to the code used to upload the same type of data in the past are needed, usually caused by the data version changes within DDR2. IFPS data will require more effort, as these are text files, which need some automatic processing. Passenger itineraries also need to be updated and loaded for September 2017. Note that in the meantime, the partners have started to work on earlier data from 2014 previously used in other projects. When the new data are loaded into the database, as these will be in the same format, the update process will be relatively seamless. In this way, data availability will not slow down the development of the rest of the project.

Access to the data for all the partners is now completely set-up, but some adjustments might be needed, for example regarding the configuration of the database. The team will also need to manage the new personnel joining the project (notably from UniBo) and any potentially leaving it.

Finally, the team will actively monitor the needs for data, should new ones arise. WP2 runs almost until the end of the project to make sure that there is no bottleneck due to data availability and that new data can be obtained and prepared, if required.

6 References

Cook, A., & Tanner, G. (2015). *European airline delay cost reference values - updated and extended values (Version 4.1).*

EUROCONTROL. (2018a, June 8). *DDR2-Webportal*. Retrieved from http://www.eurocontrol.int/articles/ddr2-web-portal.

EUROCONTROL. (2018b, June 8). *Base of Aircraft Data (BADA)*. Retrieved from http://www.eurocontrol.int/services/bada

EUROCONTROL. (2018c, June 8). *Central Office for Delay Analysis (CODA)*. Retrieved from https://www.eurocontrol.int/articles/central-office-delay-analysis-coda

7 Acronyms

ABM: Agent-based model

ACI EUROPE: Airport Council International Europe

AIRAC: Aeronautical Information Regulation and Control

ANSP: Air Navigation Service Provider

ATC: Air Traffic Control

ATFCM: Air Traffic Flow and Capacity Management

ATFM: Air Traffic Flow Management

ATM: Air traffic management

AU: Airspace user

BADA: Base of Aircraft Data

CODA: Central Office for Delay Analysis

CPU: Central Processing Unit

CRCO: Central Route Charges Office

DCI: Dynamic cost indexing

DDR2: Demand Data Repository

E-AMAN: Extended Arrival Manager

ECTL: Short name of Domino partner: EUROCONTROL

GDS: Global Distribution System

GPU: Graphical Processing Unit

IFPS: Integrated Initial Flight Plan Processing System

INX: Short name of Domino partner: Innaxis

NDA: Non-Disclosure Agreement

NM: Network Manager

PRB: Performance Review Body

SSL: Secure Sockets Layer

UDPP: User Driven Prioritisation Process

UniBo: Short name of Domino partner: Università di Bologna

UNITS: Short name of Domino partner: Università degli studi di Trieste

UoW: Short name of Domino coordinator: University of Westminster

-END OF DOCUMENT-