AICHAIN Operational Value Interim Report

Deliverable ID:	D3.1
Dissemination Level:	PU
Project Acronym:	AICHAIN
Grant:	894162
Call:	H2020-SESAR-2019-2
Topic:	SESAR-ER4-04-2019
Consortium Coordinator:	SITA EWAS
Edition Date:	8 Oct 2021
Edition:	00.02.00
Template Edition:	02.00.02

Founding Members
EUROPEAN UNION EUROCONTROL
SESAR JOINT UNDERTAKING

Authoring & Approval

Authors of the document

Name/Beneficiary	Position/Title	Date
Sergio RUIZ / EUROCONTROL	WP3 Lead	8 Oct 2021
Ramon DALMAU / EUROCONTROL	WP3 contributor	5 Oct 2021
Manuel MATEOS / NOMMON	WP3 Co-Lead	1 July 2021

Reviewers internal to the project

Name/Beneficiary	Position/Title	Date
Sergio RUIZ / EUROCONTROL	Tech & Scientific Coordinator	8 Oct 2021
Javier BUSTO / SITAEWAS	Project Coordinator	8 Oct 2021
Andre RUNGGER /SWISS	SWISS representative	6 July 2021
Manuel MATEOS / NOMMON	NOMMON representative	5 Oct 2021
Daniel / SCALEOUT	WP2 Lead	10 July 2021

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

Name/Beneficiary	Position/Title	Date
Javier BUSTO / SITAEWAS	Project Coordinator	3 June 2021
Sergio RUIZ / EUROCONTROL	Project Technical Coordinator	5 Oct 2021
Daniel ZAKRISSON/ SCALEOUT	WP2 Lead	28 May 2021
Andre RUNGGER /SWISS	SWISS representative	1 June 2021
Manuel MATEOS / NOMMON	NOMMON representative	1 June 2021

Rejected By - Representatives of beneficiaries involved in the project

Name/Beneficiary	Position/Title	Date

Document History

Edition	Date	Status	Author	Justification
00.00.01	30 June 2021	Ed 1.0	Sergio RUIZ	Final draft to partners
00.00.02	10 July 2021	Ed 1.0	Manuel MATEOS	Amendments to final draft
00.01.00	15 July 2021	Ed 1.0	Project WP3 team	1st Submission to SJU
00.02.00	5 October 2021	Ed 1.1	Sergio RUIZ	Amendments requested by SJU reviewers. Added Chapter 7.

Copyright Statement

AICHAIN

A PLATFORM FOR PRIVACY-PRESERVING FEDERATED MACHINE LEARNING USING BLOCKCHAIN TO ENABLE OPERATIONAL IMPROVEMENTS IN ATM

This Project Management Plan is part of a project that has received funding from the SESAR Joint Undertaking under grant agreement No 894162 under European Union's Horizon 2020 research and innovation programme.

Abstract

The AICHAIN Operational Value Interim Report provides the progress and outcomes from project WP3 up to the project mid-term, that is: identification of the ATM use case, model, data sets, and benefits-impact model (BIM) preliminary analysis (Task 3.1), the description of design of experiments, FedML models, and data use terms (Task3.2), and the initiated development of FedML models and data preparation (on-going Task 3.3).

The AICHAIN project is about defining and assessing the innovative concept of *privacy-preserving federated machine learning* applied to the ATM domain. The concept is proposed as a new way to exploit in a privacy-preserving and cyber-secured way the valuable information available in ATM stakeholders' private datasets for machine learning applications. Those are datasets that are currently in silos and there is reluctance and barriers preventing their exploitation.

Table of Contents

List of Tables

List of Figures

1 Introduction

1.1 About the AICHAIN project

The AICHAIN project aims at enabling the cyber-secured exploitation of large private data sets that belong to different stakeholders and that contain valuable information for ATM operations. To overcome the stakeholders' reluctance to share their sensitive data, the exploitation will not be performed by exchanging data directly but by articulating an advanced privacy-preserving federated learning architecture in which neither the training data nor the training model need to be exposed or shared. This will be possible thanks to the innovative combination of two emerging digital information management (DIM) technologies: Federated Machine Learning (FedML) and Blockchain.

In the AICHAIN project the combination of FedML and Blockchain technologies will be explored (while not excluding other privacy-preserving and cyber-securing technologies) to fully unblock the full potential of the two novel concepts. In particular:

1. The use of FedML provides a first degree of privacy since the sensitive private data never leaves the DIM infrastructure of the data owner. This allows reducing the cyber-security threats significantly, which can be addressed with typical security techniques (Firewalls, IP filters, anti-malware, among others).
2. Using Blockchain to create an immutable audit trail for the federated models to ensure the trustworthiness and integrity of the information exchanges while proving the data and model provenance.
3. Using Blockchain smart contracts and tokens for governing the business logic of the FedML alliance. Smart contracts and tokens can be used to define rules, rewards and penalties (i.e., an incentive mechanism) around an agreement, automatically enforcing those obligations. The idea is to encourage the FedML alliance members to contribute independently and be incentivized to improve the common model as much as possible.

The **specific objectives of the AICHAIN project** have been structured in three areas of research (see Figure 1-1), to be covered with different levels of deepness: i) the DIM **technological** solution, ii) the **operational value** of the DIM solution, and iii) the **governance & incentives** aspects. The following are the **specific objectives** in those three areas:

- **OBJECTIVE #1**. In the **technology dimension**, to define the **AICHAIN Solution architecture** as a potential SESAR technology enabler for the exploitation of private data value, and to implement a functional small-scale **prototype for user validation and operational value experimentation**.
- **OBJECTIVE #2**. In the **operational dimension**, to demonstrate and quantify the **operational value** of the AICHAIN concept with an **ATM use case** in the area of Advanced Demand Capacity Balancing (DCB) services.
- **OBJECTIVE #3**. In the **governance dimension**, to develop an **incentive mechanism** that addresses the motivational aspects of the data owners in order to facilitate the adoption and the effective utilisation of the AICHAIN concept.

This report D3.1 covers the Objective #2 Operational value (i.e. value for ATM operations) of the project.

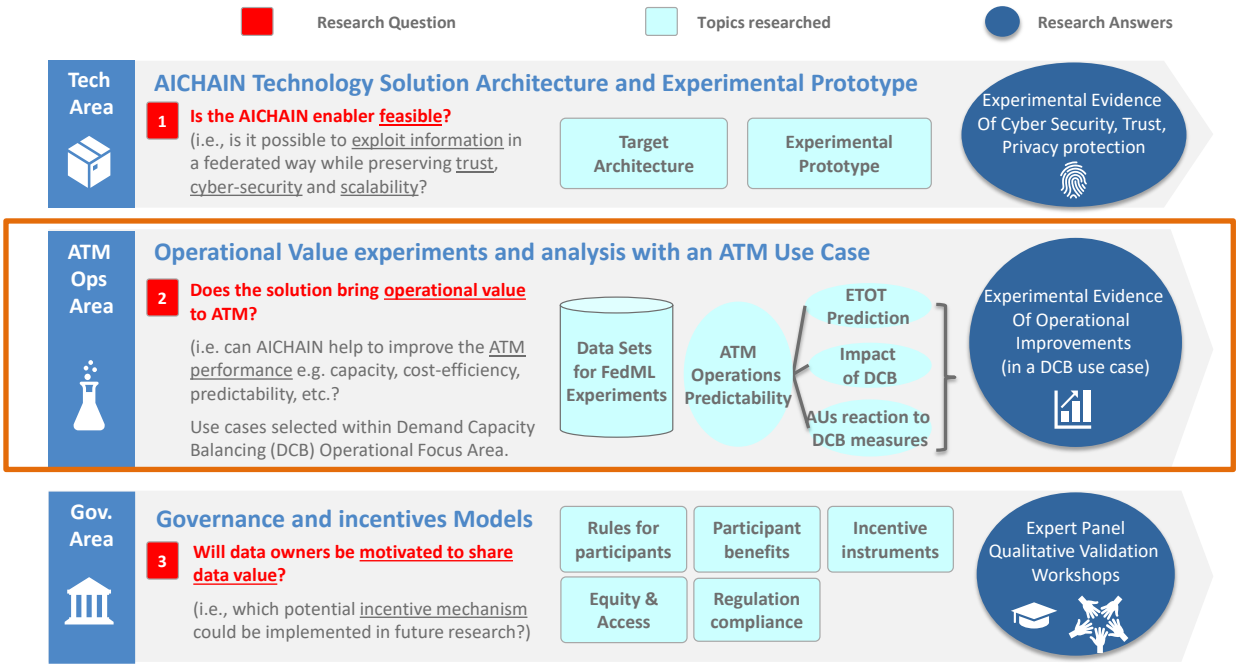


Figure 1-1 AICHAIN project research areas and questions

1.2 Purpose and scope

This deliverable document **D3.1 AICHAIN Operational Value Interim Report** provides the progress and outcomes from project WP3 up to the project mid-term, that is, the identification of the ATM use cases, models, datasets, and benefits-impact model (BIM) (Task 3.1), the description of design of experiments, FedML models, and data use terms (Task3.2), and the initiated development of Fedml models and data preparation (on-going Task 3.3).

Next Figure 1-2 recalls the overall project methodology and workflow of tasks, depicting the scope covered in the present report. The quantitative evidence of an early preliminary experiment, conducted in this first period of the project, which is beyond the due scope of this report, will be also presented and discussed in chapter 8.

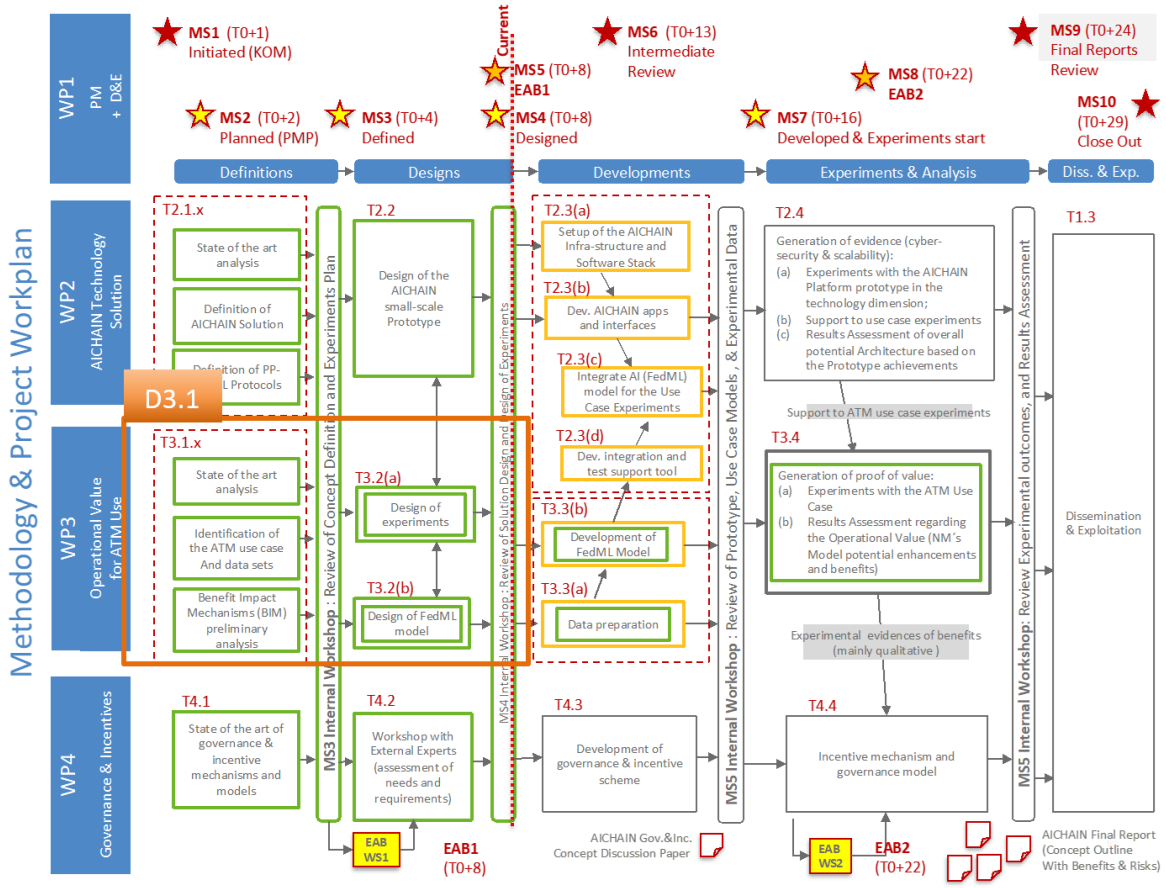


Figure 1-2 AICHAIN project methodology: work flow and tasks

To **demonstrate the value** of the AICHAIN technological solution in the ATM context, the research will explore both qualitative and quantitative evidence.

First, a qualitative analysis of the possible contributions and **relevance of AICHAIN as a potential new SESAR enabler** will be conducted.

In addition, to generate quantitative evidence the project will explore two use-cases from **Advanced Demand and Capacity Balancing (A-DCB)** domain that have been identified among the ones of highest interest and priority for the Network Manager, where the private data exploitation enabled by the AICHAIN solution could potentially bring significant operational value. Note in any case that the AICHAIN Technological Solution proposed could be useful in other contexts as well, e.g., advanced air traffic services, total airport management, or others.

The DCB use cases to be developed in this project are shortly described below and further developed in following sections of this deliverable:

1. **Prediction of expected take-off time (ETOT).** This use case is today in operations, but using NM data only. It has been found as relevant to research in which the privacy-protected

exploitation of Airspace Users' data –through the AICHAIN technological solution– could significantly improve the prediction performance of the current ML approaches.

2. **Prediction of AUs' reaction (change to flight plans) to network and weather constraints.** At present, most of the flight plans (FPL) are filled during the day of operations (during tactical ATFM phase). Nevertheless, these initial FPLs often suffer changes/updates until they become actual flight trajectories, e.g. due to the impact of severe weather conditions or unexpected changes of ATFM measures. Having a better prediction that takes into account these conditions could be a key input for FMPs and the NM to decide whether and how a network constraint (e.g., a regulation) should be activated based on the anticipation the potential reaction of AUs, which will determine at the end the actual effectiveness of such ATFM measure. In this way more efficient constraints could be designed, thus reaching more stability and predictability in the network and a better quality of service for AUs.

This deliverable reports on:

- A **discussion of why, how and for what AICHAIN is relevant to ATM**
- The **description of the two ATM-DCB use cases**;
- The **preliminary benefit-impact mechanism (BIM) of the use-cases**;
- The **design of the ML models and datasets** (including the refinement of features needed);
- The **validation plan and the design of the experiments** to be conducted; and
- The **early experimental quantitative evidence** for one of the use-cases.

1.3 List of acronyms

Table 1-1 provides the list of acronyms in used across the document, combining ATM domain and DIM (technology) domain acronyms.

Acronym	Description	Domain
A/C	Aircraft	ATM
AI	Artificial Intelligence	DIM
ANSP	Air Navigation Service Provider	ATM
AO	Airport Operator	ATM
ATC	Air Traffic Control	ATM
ATCO	ATC Operator	ATM
ATFCM	Air Traffic Flow and Capacity Management	ATM
ATM	Air Traffic Management	ATM
AU	Airspace User, e.g. an airline	ATM
BIM	Benefit Impact Mechanism	ATM
CBA	Cost Benefit Analysis	ATM
DCB	Demand Capacity Management	ATM
DIM	Digital Information Management	DIM

EAB	External Advisory Board	ATM
ETA	Expected Time of Arrival	ATM
ETFMS	EU/ECTRL Traffic Flow Management System (an ATFMS system)	ATM
FF-ICE	Flight and Flow Information for a Collaborative Environment (ICAO initiative)	ATM
FML	Federated Machine Learning (also used FedML)	DIM
GBDT	Gradient Boost Decision Tree (ML term, an algorithm)	DIM
HFML	Horizontal Federated Machine Learning (a category of FML)	DIM
IT	Information Technology	DIM
ML	Machine Learning	DIM
ML-OPS	Machine Learning Operations	DIM
NM	Network Manager	ATM
OFA	Operational Focus Area	ATM
OI	Operational Improvement	ATM
PET	Privacy Enhancing Technology	DIM
PP	Privacy-preserving	DIM
PPFML	Privacy-preserving Federated Machine Learning	DIM
SWIM	System Wide Information Management	ATM
TBO	Trajectory Based Operations	ATM
TFML	Transfer learning Federated Machine Learning (a category of FML)	DIM
TOT	Take-off-Time	ATM
TP	Trajectory Prediction	ATM
VFML	Vertical Federated Machine Learning (a category of FML)	DIM
XGBoost	eXtream Gradient Boost (ML term, an algorithm)	DIM
STACKn	Scaleout baseline cloud platform for ML, DML, FML, and PPFML	DIM
FEDn	Federated ML note (Scaleout baseline platform)	DIM

Table 1-1 List of acronyms in use

1.4 List of references

Appendix A provides the consolidated list of literature background references used along the document, in particular in the section of state of the art.

1.5 Applicable Reference material

[1] Grant Agreement No 894162 AICHAIN – Annex 1 Description of the Action.

[2] AICHAIN Consortium Agreement, Issue 1, October 2020.

[3] AICHAIN D1.1 Project Management Plan, Edition 01.00.00.

[4] AICHAIN D3.1 AICHAIN Operational Value - Interim Report, Edition 00.01.00.

[5] Project Handbook V3.0 for ER4 call:
https://stellar.sesarju.eu/jsp/project/qproject.jsp?objId=16442516.13&resetHistory=true&statInfo=Ogp&domainName=saas

[6] STELLAR Training Session Slides, 25th May 2020:
https://stellar.sesarju.eu/servlet/dl/ShowDocumentContent?doc_id=18574456.13&att=attachment&statEvent=Download

[7] Technical Specification of SESAR 2020 Exploratory Research 4 Call - ER4:

https://ec.europa.eu/research/participants/data/ref/h2020/other/call_fiches/jtis/h2020-call-doc-er4-sesar-ju_en.pdf

[8] H2020 Participant Portal Online Manual: http://ec.europa.eu/research/participants/docs/h2020-funding-guide/index_en.htm

[9] H2020 Annotated Model Grant Agreement: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/amga/h2020-amga_en.pdf

[10] SJU Model Grant Agreement: http://ec.europa.eu/research/participants/data/ref/h2020/other/mga/jtis/h2020-mga-er-sesar-ju_en.pdf

[11] OECD. *Enhancing Access to and Sharing of Data Reconciling Risks and Benefits for Data Re-use Across Societies*. OECD Publishing, 2019.

1.6 Document structure

The document is structured as follows:

- **Section 2** discusses the relevance of AICHAIN as a key enabler to support the higher levels of automation needed in the future Digital European Sky concept.
- **Section 3** elaborates the state of the art and the identification of use cases in the domains of ATM and ATFCM/DCB.
- **Section 4** describes the two use cases selected for the experimentation and elaborates their preliminary benefit impact mechanisms (BIM).
- **Section 5** is about the design of the ML model and datasets for the use-case 1 (UC1)
- **Section 6** is about the design of the ML model and datasets for the use-case 2 (UC2)
- **Section 7** presents the validation plan of the AICHAIN project, which is needed to fully understand the experimental plan presented in the next section.
- **Section 8** experimental plan and methodology (i.e. design of experiments) to generate the due evidence for the two use-cases (and more generally, about the AICHAIN solution).
- **Section 9** presents some early evidence for UC1 in which a proof of value of the private datasets is given in a non-federated setup
- **Section 10** summarises the conclusions and future steps
- The **Appendix** contains all the references used as state of the art in this document.

2 AICHAIN as one key enabler for the future Digital European Sky

2.1 The Digital European Sky

The Single European Sky (SES) initiative was launched in 2001 with the goal of improving the performance of air traffic management (ATM) in Europe in terms of increasing the ATM capacity, safety and resilience, while reducing both the cost inefficiencies and environmental impact of the air transport.

Since then, multiple studies, research programmes and development roadmaps have been developed to contribute to such modernisation of the ATM and the effective implementation of the SES. Among the most updated references, the following are particularly relevant to understand the current expected ATM evolution: Strategic Research and Innovation Agenda (SRIA), the recently adopted European ATM Master Plan edition 2020, the Airspace Architecture Study (AAS) and its AAS Transition Plan, The European Green Deal, Flightpath 2050, among others [1]-[8].

Due to the recent and strong rise of digital data technologies, e.g. high-speed and cyber-secured telecommunication networks, super-computing capabilities, big data warehouses, deep learning, etc., most of the industries are living a revolution towards what is called the *digital transformation*, in which higher levels of automation are being put in place to support more advanced planning and execution of operations. The ATM is also under such process of digital transformation, and is moving towards what has been recently labelled as the *Digital European Sky*. The aim is to make the Digital European Sky the most efficient and most environmentally friendly sky to fly in.

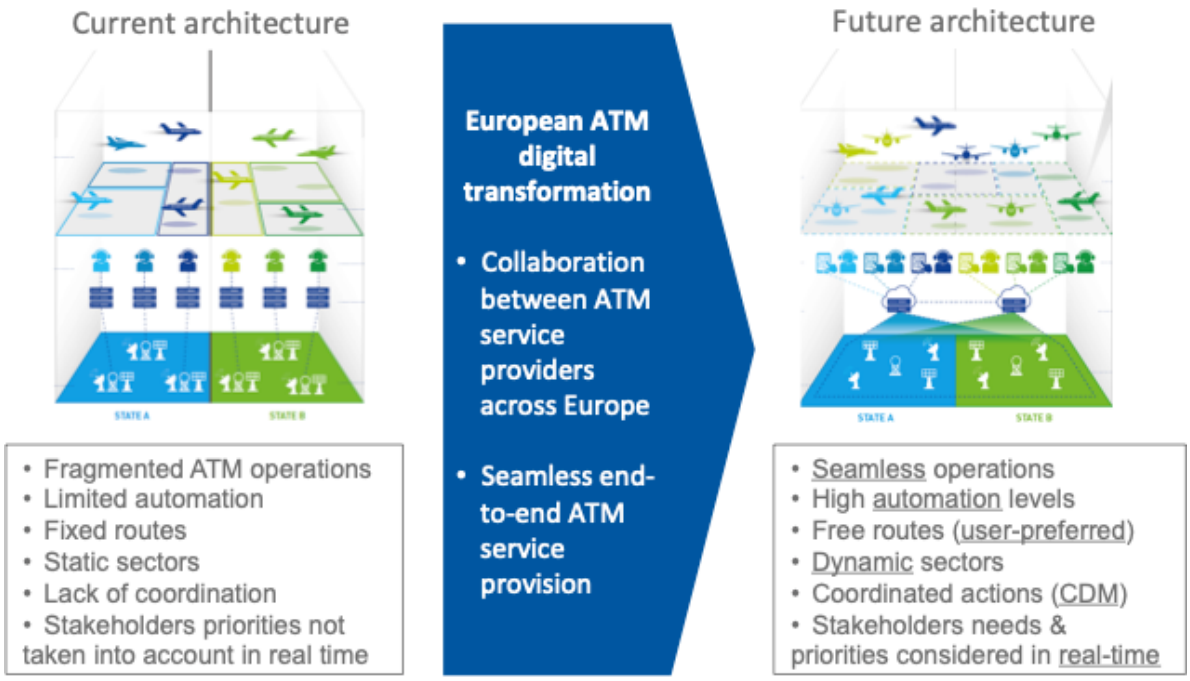


Figure 2-1 The aimed evolution of the Airspace Architecture

Figure 2-1 highlights some of the major changes targeted for the Airspace Architecture. In a nutshell, the digital transformation of the European ATM will enhance the level of coordination of the multiple different actors that operate in the complex logistical system that is the European ATM, providing a common situational awareness and facilitating the collaborative decision making (CDM) to deliver a seamless end-to-end ATM service provision. This will reduce the traditional inefficient fragmentation of the ATM operations at the maximum extent, and will enable the accommodation of the traffic demand as close to their optimum plans as possible. Flights would fly user-preferred routes and optimal vertical profiles, instead of flexed routes and inefficient vertical procedures, and the capacity at sectors will be allocated dynamically and adapted to the demand with the goal to set the minimum and lesser impacting number of ATM constraints as possible to both the flows and to the individual flights. Relying on higher levels of automation, real-time common situational awareness and advanced CDM processes, it will be possible to take into account the priorities and needs of all the relevant stakeholders (NM, ANPSs, Airports, and Airspace Users) in order to synchronise all the actors and achieve the best ATM service performance possible at any moment.

The resulting seamless operations enabled by the new Airspace Architecture described above for the Digital European Sky is called *Trajectory Based Operations* (TBO). Figure 2-2 shows a timeline view of the seamless and coordinated planning of traffic trajectories and ATM constraints. All the actors will be connected to the digital ATM eco-system through the SWIM (System Wide Information Management), and their operations will be coordinated and synchronised through collaborative planning of *4D Trajectories*. The '*4D Trajectory'* will be a digital representation of each flight plan that incorporates, with high level of detail, the intended spatio-temporal evolution of each flight.

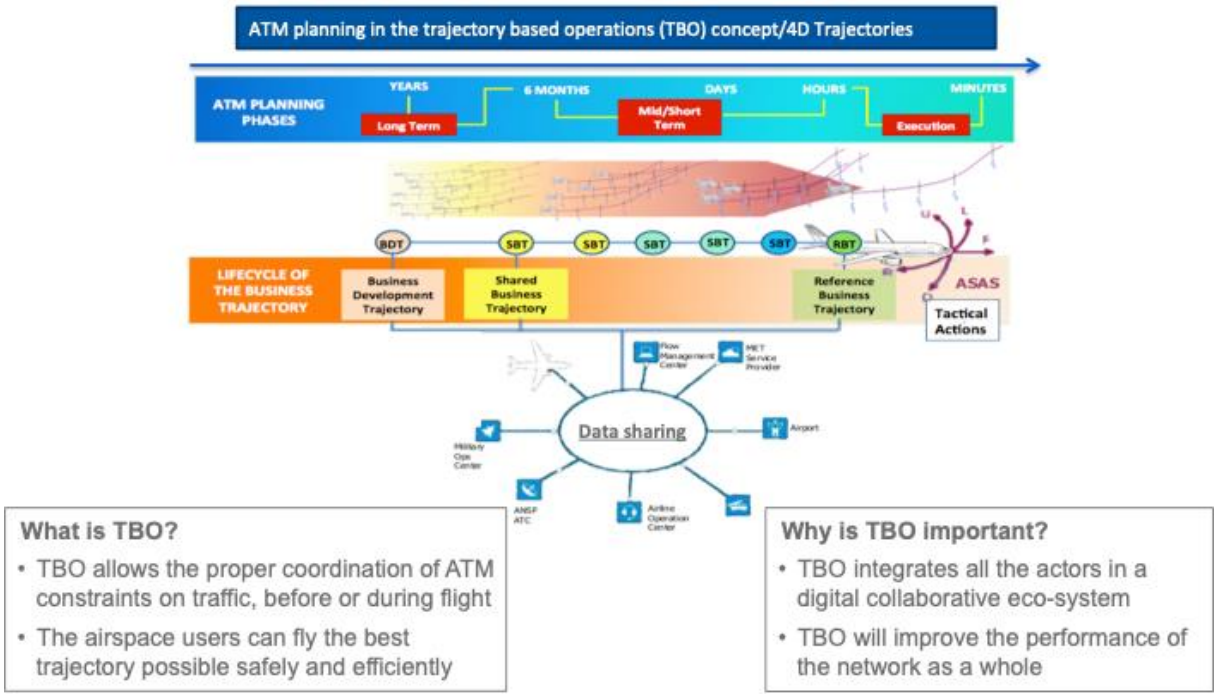


Figure 2-2 The concept of Trajectory Based Operations at a glance

Due to the presence of uncertainties, each flight will indeed have multiple alternate feasible 4D Trajectories that will evolve at the different planning phases. The process of 4D Trajectory development will start internally by the airspace users, and is called the Business Development Trajectory (BDT). At some point in time, the trajectory will be shared with the digital ATM eco-system, thus becoming the Shared Business Trajectory (SBT), which will become the main digital flight plan model enabling the collaborative planning among all the relevant ATM actors. The SBTs will evolve and become more concrete as soon as the time gets closer to the time of flight execution and the uncertainties are reduced. At some moment in time, estimated around 15-20 minutes before the take-off time, the SBT will be declared as RBT (Reference Business Trajectory) and will become the plan that the airspace user is compromised to fly and the ATM service is agreed to facilitate.

2.2 The operational value of AICHAIN in the Digital European Sky context

It is worth noting that the target TBO concept has traditionally relied in the principles of data sharing. However, in this context **data privacy is a major challenge**: some relevant pieces of **operational information are strategic and sensitive for ATM stakeholders** (e.g. cost structure of flights, fuelling policies, aircraft weight, aircraft and crew schedules, passengers' connectivity, among others). Thus, ATM stakeholders are **highly reluctant to disclose business-sensitive information,** like for instance their business preferences, priorities and procedures. In addition, **data protection laws** (i.e. GDPR) also impose some privacy constraints to data sharing (e.g., passenger data, since involving personal information, must be highly protected). Due to all that, some of the private data that is relevant to achieve an efficient and resilient ATM planning is indeed ***non-shareable data***.

In contrast, the digitalisation of the ATM requires more and more data. In [5] the long-term vision of the automation in ATM is analysed, and it is stated that the rise of new artificial intelligence (AI), and more in particular the so-called **machine learning (ML) and deep learning (DL) methods, can provide real opportunities for a fundamental change in the automation landscape of the ATM**, which is fundamental to support the future Airspace Architecture and TBO operations. But all these **ML/DL techniques require the exploitation of data-derived information**, rather than formalised human knowledge, which **an important share of it will be non-sharable data** as discussed above.

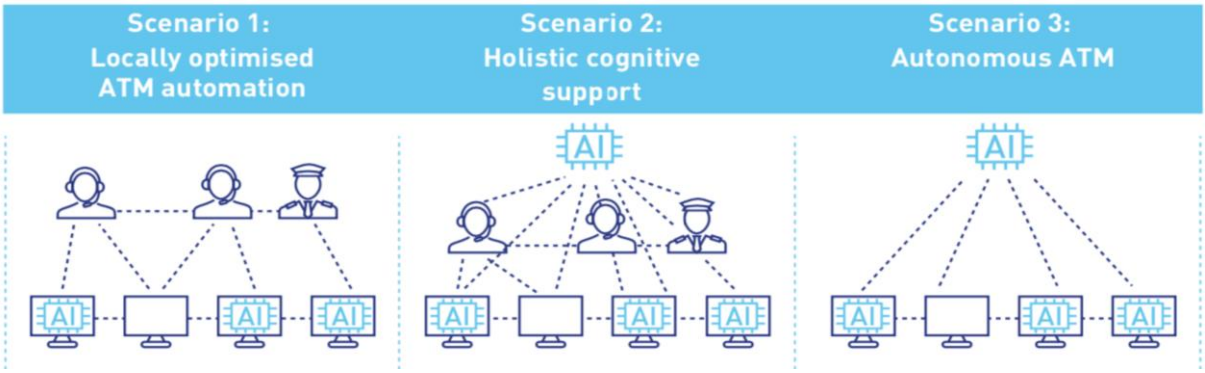


Figure 2-3 Future automation scenarios as candidates for a long-term vision for automation in ATM

Figure 2-3 illustrates the three main automation scenarios considered so far in such long-term vision [5]. The scenario 1 (S1) is about keeping the human as the main and only decision-maker of the core ATM processes and use the automation supported by AI models (mostly composed by expert systems and machine/deep learning models). No major paradigm shift is expected in such scenario with regards the ATM ways of working, just some enhancements that would evolve locally to optimise the awareness and productivity of human operators in their current tasks. The scenario 2 (S2) is a system based on S1 in which the human remains in control, but with some of the automation capabilities capable of initiating actions and of making decisions that could even override human decisions in some cases. Information workflows would be adapted to the human processing capabilities and human tasks would be monitored and enhanced by AI models to reduce human-related errors and keep their stress, productivity and performance at optimum levels. Lastly, scenario 3 (S3) counts with a fully autonomous ATM system in which the AI controls all the processes and the human intervention is non-existing in practice. In this scenario the automation can fully substitute the human and the system is not limited anymore by human workload constraints thus being much more scalable and faster. In terms of roadmap evolutions, the S1 is the one expected to be addressed in the short and mid terms. However, it is expected that such scenario cannot cope with the foreseen traffic levels of the next decades. S2 and S3 should be able to handle the forecasted increase in traffic demand, but it will require much more advanced AI techniques to compensate the human resilience and capabilities to adapt to unknown/unseen situations, which is considered too disruptive and risky at this moment, taking into consideration the low level of maturity of the current AI technologies and their projection in the next decades. The analysis concludes that the “holistic cognitive support” scenario (S2) should be pursued in the mid (2035) to long-term (2050) future in ATM, while S3 could be an option to consider for U-space, with a later potential transfer and adoption by ATM.

Some specific ML/DL applications in ATM as soon as the levels of automation are increased could be:

- **Trajectory based operations (TBO):** take-off time prediction, trajectory prediction, traffic demand prediction, trajectory planning support, and others.
- **Enable new airborne capabilities:** 4D trajectory precise execution, airborne self-separation, drones & U-space, and others.
- **Boosting airline performance:** minimise the impact of ATM constraints, optimal 4D Trajectory planning, reduction of the environmental impact, and others.

On the other hand, the computational burden required to train complex advanced AI/ML models may limit them to a reduced number of applications. Due to the large number of stakeholders and flight operations per day, **the training of deep learning models for real-time ATM applications could require the parallelisation/de-centralisation of the training process, e.g. for real-time applications**.

Putting all together, **the lack of transparent data access to relevant operational private data and the high performance computation requirements undermine today the possibility of taking full advantage of ML/DL techniques in ATM** to improve the predictability of the operations and ultimately to optimise and improve the quality of service. Is in that context where the AICHAIN solution could generate significant value for the ATM, i.e. to support the increasing automation levels, to enable increased predictability, to enhance the common situational awareness, and to enable seamless collaborative decision making between multiple actors for planning optimisation. Figure 2-4 summarises at a glance the main value drivers of the AICHAIN Technological Solution proposed, which should have value in the S1 automation scenario (for today’s applications) but even more in the foreseen long-term S2 and S3 scenarios.

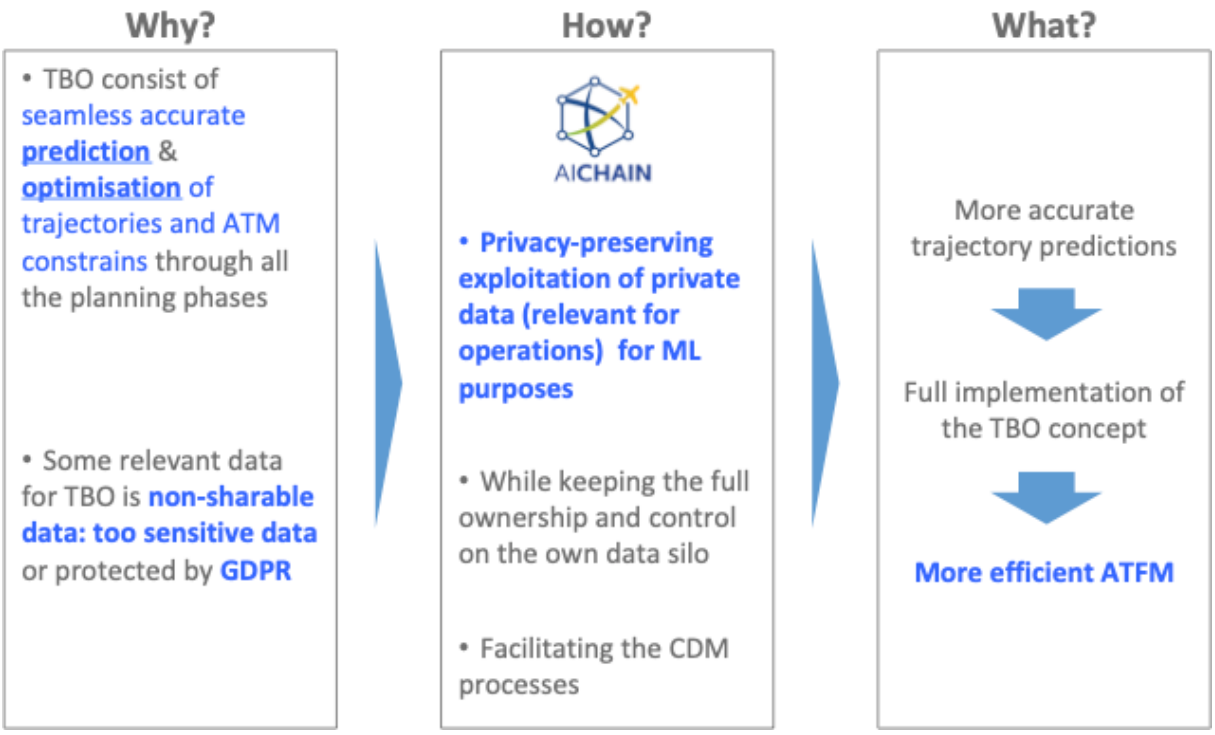


Figure 2-4 The operational value of AICHAIN at a glance

2.3 Link to the FLY AI Report recommendations

The FLY AI Report [7] aims at giving support to the realisation of the Digital European Sky and the SESAR target vision. The report acknowledges that increased levels of automation, cyber-secure data-sharing and connectivity are key to achieve the required digital transformation and to support the envisioned concept of operations. And AI is identified as a key technological enabler to develop the desired higher levels of automation in ATM.

The report sets up a series of recommendations among which the first one of them is to establish a federated data foundation and AI-infrastructure to give support to the development AI-based solutions, while respecting the following key principles that are also taken into account in AICHAIN:

- The collection by an ATM data community of all available operational data in the federated **network of data lakes**, the **nomination of federators** in charge of collecting and curating data, and an easy access to **computing power** through the AI-infrastructure. All these three points are addressed by the federated and de-centralised architecture proposed in the WP2 of the AICHAIN project (see Deliverable D2.1). The AICHAIN solutions will offer secure collecting, secure storing, managing and maintaining cyber-security data sets for Aviation/ATM AI-based developments in a systematic, harmonized, processable and decentralised way.

- The **sharing of data** whilst respecting the following principles: self-determined **control** over stored and processed data and decision on **who is permitted** to have access to data. In the AICHAIN solution, rather than sharing raw private data, what is indeed shared are the encrypted ML model contributions. This is one of the techniques used to keep the privacy

and control of the data as required by the data owners. Deliverable D2.1 describes in detail the proposed solution to cover the requirements of privacy and control of the federated private datasets.

- The archiving of all **real-time** data to support the training phase of AI/machine learning developments. This is also addressed in the design of the AICHAIN solution, and it is part of the so-called ML-OPS capabilities to collect and process all operational data in a continuous and manner and well adapted to the required operational lead-times of each application.

- The use of **open data exchange standards**. The AICHAIN is proposed as an open architecture and the software developed to build the AICHAIN prototype is open-source code.

- The establishment of a **data governance** to ensure that the community is collecting, storing and using data appropriately to create shared value. An indicator of data quality and model contribution merits should be introduced to clarify who is responsible for what during the data production, dissemination and storage and who is controlling whom and who is liable for what along the entire data cycle. This point is addressed in WP4 of the project (see Deliverable 4.1).

In conclusion, the AICHAIN Technological Solution is well aligned to these recommendations and therefore it can be considered as a qualitative evidence of the potential of AICHAIN to generate value for ATM and its potential to become a key enabler of the future Digital European Sky.

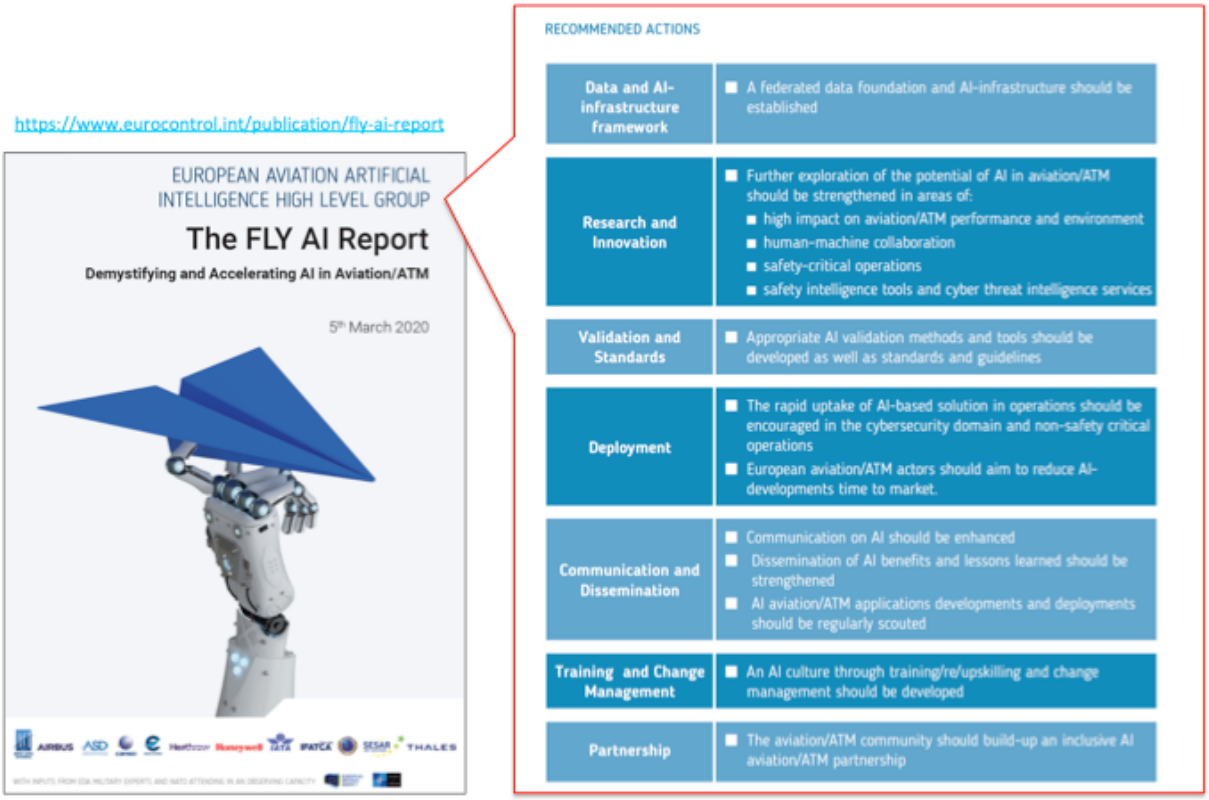


Figure 2-5 The ten recommendations of The FLY AI report

2.4 AI/ML applications for ATM identified in the FLY AI Report

The FLY AI Report [7] identified eight key areas in which AI-based applications are expected to generate value in the context of aviation/ATM. Table 2-1 shows the identified eight areas and some of the expected benefits that could be achieved in each area.

Table 2-1 Aims and benefits of the AI applications of the catalogue (source: FLY AI Report)

AREA	BENEFITS
Traffic predictions / forecasts/modeling	1. Improving predictions of aircraft trajectories, reducing uncertainty and increasing capacity
Resource management / Optimisation	1. Deploying the optimal configuration of sectors and thus optimising capacity with the available resources 2. Supporting ATM demand and capacity balancing
Workload / Automation / Autonomy	1. Reducing ATCO workload (e.g. using speech recognition models for controller assistance) 2. Reducing risks with safety intelligence tools[11]
Airport performance	1. Improving runway throughput (e.g. ROT prediction, improving spacing buffers) 2. Cutting airport delays
Passenger experience	1. Improving passenger transfer/ customer satisfaction 2. Using biometrics to accelerate secure boarding
Infrastructure monitoring	1. Improving GNSS monitoring 2. Cybersecurity monitoring
Airborne capabilities	1. Improving validation capabilities 2. Generating environmental improvements 3. Pilot and ATCO assistant through automatic speech recognition 4. Enhancing safety with automatic taxi, take-off and landing enabled by computer vision
Airline performance	1. Optimising fuel usage 2. Proposing better and more routes

In addition to these areas, the FLY AI Report also identifies more than 20 use-case applications in the aviation/ATM context, in which different AI/ML models have been developed and sometimes used in operations. Figure 2-6 shows a map of these identified AI/ML applications in which they are sorted according to the their technological and operational maturity/readiness.

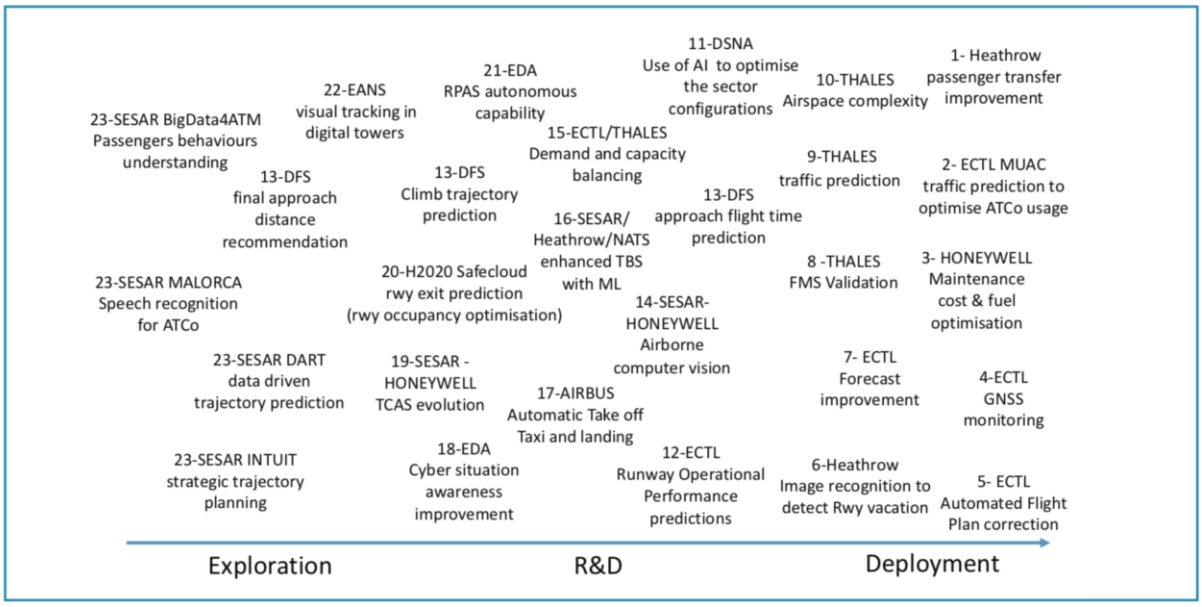


Figure 2-6 Maturity assessment of selected aviation/ATM AI applications (source: FLY AI Report)

2.5 Some areas of the ATFCM/DCB function that could benefit from federated ML

The ATFCM function is all about predicting and controlling/mitigating the predicted risks of aircraft separation losses and ATC workloads of the future network estates. But good predictions need good data. But an important part of this data is owned and kept private by the airspace users, thus hindering the options for the NM and ANSPs to make the best predictions possible and, in consequence, deliver the most effective and well-performing service. Some of the ATFCM tasks that could benefit from higher predictability capabilities are:

1. Traffic demand predictions (flight trajectories and/or flows):
    - Today more than 90% of the flight plans are sent very late during the day of operations (i.e. from 6 to 3 hours in advance to the filed departure time). This obliges the ATFCM strategic and pre-tactical functions to using flight plan predictions based on historical records. Better predictions in this area could impact positively in the processes to optimise the ANSPs sector opening scheme and regulations scheme.
2. Traffic demand updates/reactions (flight trajectories and/or flows):
    - Due to the dynamicity and uncertainty present in the network plan, the AUs often react to the new DCB and weather constraints and they often update their flight plans to adapt to the current and foreseen situation in the network. In order to propose a more effective and efficient scheme of ATFCM constraints, the NM functions could benefit from a better predictability of the AUs reactions to those constraints. This problem can be found either at pre-tactical or tactical phases, but it becomes more important during tactical ATFCM because the levels of uncertainty of

the network and weather constraints are much lower and the level of reactivity to those constraints is much higher.

3. Entry counts and occupancy counts:

   - As soon as the traffic demand gets closer to execution is more and more important to have a good prediction of the entry counts at each sector, to be able to mitigate any risk of potential ATC overload while avoiding overprotection of the ATC sectors. Even assuming that the flight plans are stable at that stage (e.g. less than 2 hours from execution) there is still an important degree of uncertainty in the traffic counts that may produce important variations in the times at when the flights actually depart with respect the times filed in the flight plans. Therefore, having a good predictability of the expected take-off time (ETOT) is key to assess the probability of ATC overload, which will in turn determine which ATFCM constraints, if any, must be applied. This is therefore of vital importance for those sectors that are at or close to their capacity limits. And for the most complex and congested sectors, a predictability of the ETOT with absolute errors below 5 minutes could enable a better management of the ***traffic complexity***, which is directly related to capacity, i.e., the ATC workload, and necessary to optimise the number and type of ATFCM constraints to flights. See section 2.3 for more detailed state of the art on this topic

4. Selection of the best / most optimal ATFCM constraints (network optimisation):

   - The application of any ATFCM constraint (i.e. regulation or STAM) may have different impact on the flight costs and schedules of the different AUs. Today, some measures are applied with little awareness of the impact generated to AUs (e.g. regulations), or by considering assumptions that could be often not very accurate, e.g. assuming that the AU that operates that flight would prefer a STAM in which a longer route is given to avoid 30 minutes of delay in a regulation; it may happen that for this specific flight the delay is less costly than the longer route. Having a better prediction of the impact of ATCFM measures on AUs could help the NM controllers and FMPs to define a more effective and efficient set of ATFCM measures. Also, as commented in problem number 2, the AUs may react to those constraints that cause them large disruptions and costs in their operations, therefore, having a good prediction of the impact of constraints on AUs could also contribute to anticipate the potential reaction of demand. This is also a way to incorporate the priorities and preferences of AUs in a dynamic way and with little burden for the flight dispatchers.

In the AICHAIN project two use cases belonging to the first area (i.e. traffic predictions/forecast/modelling) will be analysed, since they are of high interest for the Network Manager and its main operational function of supporting Demand and Capacity Balancing in the network. One of the models (the ETOT Prediction) is a simplification of the one developed by EUROCONTROL and deployed in MUAC (see Figure 2-6), and the other (the Pre-tactical demand prediction) is an on-going R&D project developed by NOMMON in collaboration with EUROCONTROL. The description of these two cases will follow the next state of the art chapter of AI/ML in the Air Traffic Flow & Capacity Management context.

3 State of the art of Air Traffic Flow and Capacity Management and AI/ML

3.1 Air Traffic Flow & Capacity Management

3.1.1 The importance of predictability in current operations

The air traffic flow and capacity management (ATFCM) service is provided by the Network Manager Operations Centre (NMOC) to the airspace users throughout the European Civil Aviation Conference (ECAC) states (presently 44 states), with the purpose to utilize the available airspace capacity in the most cost-efficient way possible, while enabling safe, orderly and expeditious Traffic Flows. A *traffic flow* is composed by several flights moving through a given airspace region at a given time period and in a common direction

Nowadays, the key process of the ATFCM in Europe is the Demand and Capacity Balancing (DCB) [30]. The main goal of the DCB service is to ensure that airspace capacity and traffic demand match in order to avoid (unsafe) overloaded sectors at any time, i.e. demand and capacity balancing.

To do that, the ATFCM service – through the Enhanced Tactical Flow Management System (ETFMS) – makes a **prediction of the airspace demand** by computing (through roughly accurate models) the expected trajectories and their evolution over the time from the information of each individual flight plan. Also from the pre-declared information of the ATC operators it is possible to make a **prediction of the expected capacity** that will be available at every airspace sector. Those predictions are refined as the day of operations becomes closer, when the quantity and quality of information used for predictions usually increases.

DCB presents three levels of decision-making processes that vary based on the look-ahead times of the ATFCM planning, i.e. **strategic** (from 1 year up to 1 week before the day of operations D), **pre-tactical** (from 1 week to 1 day before D) and **tactical** (during all day D).

In case that any imbalance (a.k.a. *hotspot*) is detected at day of operations D or at D-1 between the predicted traffic and the available network capacity, the ATFCM shall apply ***ATFCM regulations,*** which may generate important ATCFM delays to some flights, or ***ATFCM scenarios***, which often consist in applying re-routings and/or flight level changes to some flights.

These techniques mitigate the risk of hotspots in the network since they reduce the traffic density and complexity of the traffic received by the ATC officers. However, the high number of ATFCM regulations and ATFCM scenarios in the last decade has been a major issue in the ATM, due to the **high costs supported by the airspace users (AUs)** that affects to the competitiveness of the European air traffic system [31]. This is a consequence of the lack of operational capacity compared to the existing increasing demand of traffic.

Due to the lack of predictability, the capacity estimated and declared by the ANSPs to the NM is today subject to a lot of uncertainty. Due to the presence of uncertainty (i.e. lack of predictability) ac conservative capacity declaration is made by adding relatively **large safety buffers** to the maximum number of flights that should be allowed in a sector at the same time. These buffers are necessary to

ensure that the workload of ATCOs is under acceptable levels at any moment for all the likely traffic scenarios [28].

Therefore, **higher predictability in the ATM operations, especially of the traffic flows and trajectories, could contribute to increase the capacities used in operations**, e.g. by reducing the capacity buffers at sectors. This in turn could also lead to more stable, safe and cost-effective network plans while reducing the strategic and tactical costs borne by the AUs [26][29].

3.1.2 The importance of predictability in the SESAR ATM target concept

One of the main objectives of SESAR is to facilitate a large increase of the European ATM capacity (at least a two-fold increase) by 2025+ horizons, in which the forecasted demand should be ideally allocated with minimum deviation with regards the AUs and passengers demand. To achieve these concepts the SESAR ATM paradigm introduces the concept of **trajectory-based operations (TBO),** which is aimed **to increase the predictability** of the operations and in turn to increase the capacity and the efficiency of the ATM system.

The final goal of increasing traffic predictability through the management of 4D trajectories is to better integrate the planning of the network and traffic flows with the execution of the operations, thus enhancing the performance of the ATM system in terms of safety, capacity and flight efficiency, among other performance areas. The Network Manager (NM) can facilitate the dialogue between the key stakeholders to **resolve demand/capacity imbalances in a collaborative manner** (except in timely-critical situations).

Figure 3-1 show the current ATFCM model in which potential ATC overloads (i.e. periods in which ATC officers may receive an unsafe number of traffics) are protected through the management of **dynamic density, as a proxy** to reduce the probability of bunching and the number of conflicts (since the number of conflicts grows proportionally to the traffic density in an airspace volume [32]). Traditionally, the most straightforward way to control traffic density from one day before and during the same day of operations has been the activation of regulations. In the recent years, regulations are substituted or complemented whenever possible by short-term ATFM measures (STAM) that consists of a combination of measures (re-routing, FL capping, speed changes, cherry picking delay, etc.) that are applied to some selected flights to optimise the results and minimise the impact on AUs.

To generate the extra capacity needed, the SESAR concept pivots around the idea of trajectory management at early stages of the traffic planning and execution. Figure 3-2 shows ATFCM model as foreseen by the SESAR concept definition, in which the network & local ATFCM functions are better integrated with the ATC functions through the 4D trajectory element. This function that fills the gap between ATFCM and ATC is called **integrated network and ATC planning (INAP)** [32]. Under this new paradigm a more refined set of ATFCM techniques could be applied. In particular, conflicts could be managed by an extender ATC planner (EAP) much before they could even be perceived by the ATCOs (interaction management), in a way that the number of conflicts to be solved by ATCOs can be significantly reduced (or potentially eliminated) while there is enough time to define in a collaborative way the best plan for each particular flight. In addition, the risk of ATC overload can better controlled by predicting and managing the traffic complexity expected in a look-ahead time of around 2 hours at each sector or traffic volume.

Enhancing predictability in different ATFCM processes where traffic predictions are needed is paramount to build the SESAR ATM concept and therefore to increase capacity, safety, cost-efficiency, and sustainability.

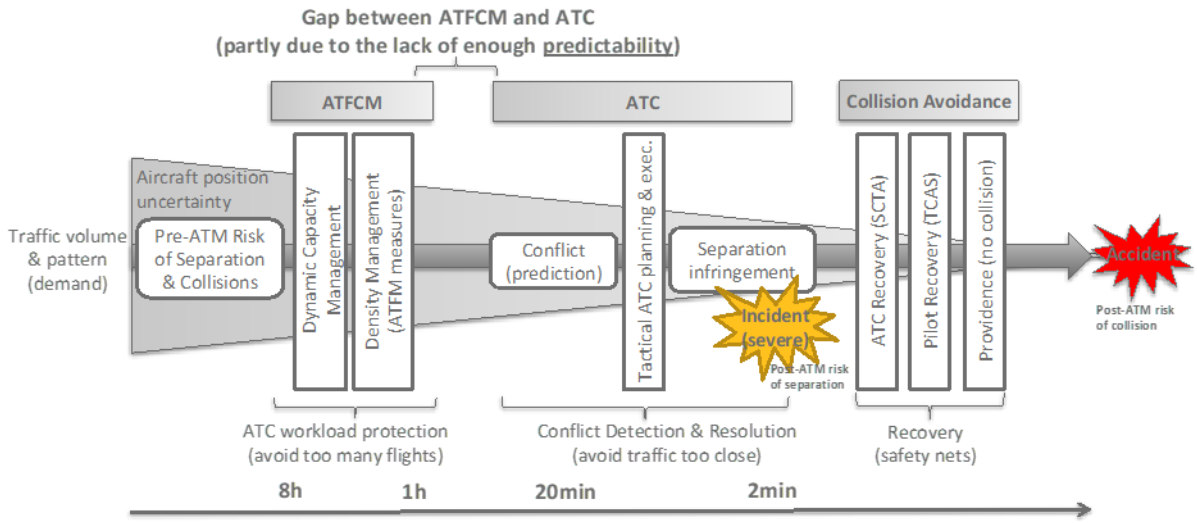


Figure 3-1 Current ATFCM/DCB model

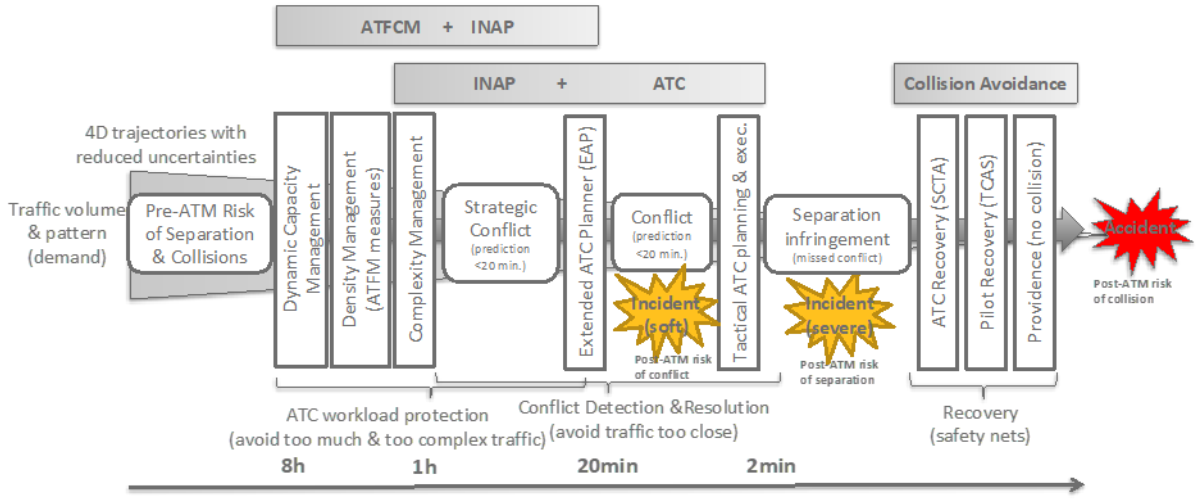


Figure 3-2 Advanced ATFCM/DCB model (SESAR target model)

3.2 Artificial intelligence and machine learning in Air Traffic Flow & Capacity Management

The aviation industry is looking at ML for a variety of different use cases, but it tends to focus on the same goals; better customer experience, improved operations and better financial results. AI-powered opportunities typically fall into one of the following five categories:

1. Predictions: Every problem where estimating future conditions can yield valuable information.
2. Optimization: Selection of a best solution or element, with regards to some criterion, from some set of available alternatives.
3. Recommendation: If there is a difficult choice and the uncertainty is high.
4. Categorization: Every instance where multiple kinds of observations need to be told apart.
5. Anomaly detection: For near real-time analysis of data flows to identify unexpected behaviour.

An important goal of the ATFCM function is to predict the demand with the greatest accuracy possible so capacity could be adjusted. To this end, two different AI approaches can be distinguished: **prediction of individual trajectories** and **prediction of the demand** (macro-model of aggregated trajectories).

3.2.1 Data driven models for trajectory prediction

Data driven models rely on the analysis of available historical data to identify the relevant state variables (input and output) and the relationships between them that are responsible for the behaviour of a system. This knowledge allows the simulation of the system behaviour under certain conditions, without knowing the physical laws that govern it. The different techniques used to generate these data driven models are grouped under the umbrella of machine learning.

Machine learning techniques can be divided into supervised, unsupervised and reinforcement learning. Supervised learning techniques (e.g. linear regression) aim to build a model from known historical data, inputs and outputs, in order to be able to foresee which will be the outputs given the inputs. Unsupervised learning techniques only use inputs without labelled responses and try to find patterns in the data; the most usual example is data clustering. Finally, reinforcement learning techniques are based on the use of agents, interacting with a certain environment, trying to maximise a reward by taking some actions; this is the approach used, for example, in some robots whose aim is to beat a human or another robot in games like chess.

Different machine learning techniques have been proposed to predict trajectories based on historical data. Since there is a large number of trajectories connecting an origin and a destination, many of them being very similar. In this kind of approach, trajectories are usually clustered to generate a discrete set of choices, each one of them represented by the average trajectory. For this reason, in this section we start by reviewing trajectory clustering techniques. Then, we present an overview of the application of data science techniques for trajectory prediction.

3.2.1.1 Clustering

Cluster analysis or clustering is the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar (in some sense) to each other than to those in other

groups. Trajectory clustering algorithms comprehend three main elements: (i) the distance metric that determines the similarity of trajectories; (ii) the variables used for the clustering, and (iii) the techniques employed to aggregate trajectories together taking as input their distances.

The most common distance metrics are those based on Euclidean distance. Since trajectories tend to present different lengths and durations, the use of the Euclidean distance requires some sort of normalisation, like trajectory point downsampling, to be a feasible comparison metric. In the case of trajectories there are other distance metrics that have been proved useful (See [13] and [14])

Regarding clustering variables, many approaches use the trajectory geometry (2D, 3D or 4D), but others include other trajectory-related features called "thematic attributes. Clustering techniques may be classified into four broad types: (i) hierarchical clustering, (ii) centroid-based clustering, (iii) distribution-based clustering and (iv) density-based clustering. A detailed review of clustering techniques can be found in [11].

As mentioned above, most applications are only based on the geometry of the trajectories. For instance [19] use DBSCAN to cluster trajectories based on their 4D trajectory geometry. To deal with Euclidean distance, the authors rely on Principal Component Analysis (PCA) over trajectories that contain up to 100 interpolated points. This combination of PCA and DBSCAN is also used in[23], in this case applied to the clustering of trajectories in the Terminal Manoeuvre Area (TMA) to predict the Estimated Time of Arrival (ETA) of a given flight. In [10] a k-Nearest Neighbours (k-NN) algorithm is used to cluster trajectory data using Dynamic Time Warping (DTW) to normalise the Euclidean distance. In [12], the authors repeat the experiment focusing on the climb phase.

Some studies have explored the clustering of trajectories based on the 4D trajectory and further thematic attributes. This is the case in the work by [16], where the authors extend the 4D domain by taking into account features like calendar properties, weather and aircraft characteristics. The distance applied to these so-called "enriched points" is also Euclidean and the method for clustering is the K-NN. In [20] the authors applied the DBSCAN technique to cluster routes (i.e., horizontal 2D projections of the trajectory) based on certain route characteristics, such as the distance travelled in each sector and the total route charges, using the Euclidean norm to measure proximity between routes.

Finally, some authors have explored the use of interactive visualisation techniques to support cluster analyses. A relevant example is the work of [9], which proposes an analytical workflow in which interactive filtering tools are used to attach relevance flags to elements of trajectories. This way, it is demonstrated how the proposed method can be used in combination with different clustering techniques and distance metrics to discard irrelevant elements of the trajectories and improve the relevance of the resulting clusters.

3.2.1.2 Predictive models

The key elements in data-driven predictive models are the techniques applied and the variables selected as predictors. The models found in the reviewed literature can be split in two main groups depending on the expected output: (i) models that aim to predict a complete 4D trajectory and (ii) models that target only the route (the 2D projection of the trajectory).

Different authors have explored the use of data-driven models for 4D trajectory prediction. In [16] the aim is to predict the 4D trajectory based on the information contained in the FPL. For this purpose, they start by clustering the trajectories of one month of flights in the Spanish airspace.

Then, they use Hidden Markov Models (HMM) to select the most probable 4D trajectory from the clustered trajectories based on the information included in the FPL. HMM models are also used in [10],[12]. In these cases, HMM are fed into the Viterbi algorithm to find the most likely sequence of hidden states. The experimentation is performed using actual trajectory data of only one flight code covering the route Atlanta-Miami over a period of 5 years. The probability of observing a certain sequence depends on the weather, in particular temperature, wind speed, wind direction, and humidity.

Other studies have focused on the prediction of the route. In [20] two approaches are compared based on multinomial regression and decision trees to predict route selection using historical data of different AIRAC cycles. To assign the most probable route to a particular flight, models choose between a discrete number of clustered routes. Initially, flights are segmented according to airline type and arrival time. Then, different machine learning techniques are explored to calculate the probability of choosing each of the clustered routes according to navigation charges, route length and the percentage of regulated flights in each one of the clusters (as a proxy of the expected congestion). In the case studies performed over three different origin-destination (OD) pairs, multinomial regression methods provided better performance than decision trees.

The influence of route charges for route selection had already been investigated in a previous work by [17]. In this work, the authors compare the cost of the routes submitted by airlines (taking into consideration charges and fuel consumption) with the cost of the shortest available route for each flight in that day. They found that for some areas of the European airspace, airlines choose longer routes with lower charges as long as this choice reduces the total cost (fuel plus charges). The authors observe that actually flown routes are usually shorter than the ones submitted in the FPL. For those routes where the extra cost of fuel associated with choosing a longer route is comparable with the charges savings, strategies of speed variations to maximise the benefits are observed.

The impact of weather on the selection of the route is addressed in [19]. The authors present the results of the route prediction in five OD pairs using four different techniques: logistic regression, support vector machines (SVM), random forests and gradient boosting. They consider the influence of 17 variables, including season, time, miles in trail and several weather-related variables. An exhaustive analysis of the results for each technique and OD pair combination is presented, showing that, while the four techniques seem to achieve reasonable performance, random forests are slightly better. The authors observe that the most relevant variables are wind, thunderstorms and rain, followed by the miles in trail.

A different approach to route prediction can be found in [21]. This research is limited to the MUAC area of responsibility, so only the portion of the routes over this area are considered. Routes are simplified using the Douglas-Pecker algorithm that simplifies the route into only four points while keeping most of the information. The trajectory prediction is performed with a deep neural network taking as input a set of heterogeneous input features, such as the Entry Coordination Point (NCOP), the Requested Flight Level (RFL), the day of the week or the reservation of military areas. The authors conclude that the proposed solution produces flight route predictions that are substantially more accurate than methods in use today. In terms of future research, they point towards the prediction of the flight ascent/descent profile and the prediction of airspace entry times at longer look-ahead horizons.

Some methods target a single dimension of trajectory prediction. For instance, [23] attempt to predict the estimated time of arrival (ETA) in the Beijing International Airport. To this end, the

authors cluster the trajectories within the TMA, and use a neural network to predict the trajectory as a function of the entry point, heading and airspeed. Reference [18] compares the ability of regression methods against the point-mass model for short-term predictions during the climbing phase. To predict the altitude of climbing at a given point they consider a set of 79 explanatory variables from radar and meteorological records that get compressed using PCA. Experiments are undertaken using real data from 1,500 climbing paths from Paris airports and linear regression, neural networks and polynomial regression. While models achieve significantly better performance using regression models, the point-mass relies on BADA parameters (standard mass and thrust settings) and improves when more accurate values of mass and thrust are considered.

Last but not least, it should be noted that one of the key reasons of low predictability in aviation is the uncertainty in departure times (i.e., the temporal dimension of the 4D trajectory). The departure time prediction problem can be effectively addressed as a delay prediction problem. The predicted departure time can be obtained by simply adding the predicted delay to the scheduled departure time. In the recent years, the use of historical data and machine learning has shown potential to improve the accuracy of delay predictions. For example the paper [33] assessed the strategic flight schedules (i.e., slots) of an airport by using machine learning to predict potential flight delays and cancellations. Another example can be found in [34], which recently proposed a causal machine learning algorithm to predict the probability (risk) of flight delays as well as to identify the driving features. In [35], an explainable machine learning model build on gradient boosted decision trees (GBDT) was used to improve the predictability of departure times. These results suggested that data-driven models could capture reactionary delays caused by delay aggregation and propagation throughout the consecutive flights of a given aircraft. This hypothesis was further reinforced in a recent publication in which recurrent neural networks were used to model the delay propagation along the sequence of flight legs operated by the same aircraft [36].

3.2.2 Demand prediction

Air traffic demand refers to the use that AUs want to make of the airspace. The most commonly used metrics are entry counts and complexity. Entry counts measure the number of aircrafts present inside a delimited volume during a certain period of time. Traffic complexity metrics aim to measure the difficulty of managing the air traffic of a given airspace at a given time period of time from the ATC perspective. Unlike entry counts, complexity is a concept without an unambiguous mathematical formulation which depends not only on the number of aircraft in the given airspace but also on the variety of directions, speeds, etc. of all the aircraft trajectories.

Air traffic demand prediction is usually performed by aggregating individually predicted or synthetically generated trajectories. An example can be found in [16], where the entry counts at different sectors were calculated as the number of trajectories intersecting the sectors boundaries at a certain period of time. There are different aspects that may affect the certainty of the estimated entry counts. In [22], the authors propose a methodology to investigate the impact of weather on the entry time uncertainty. The proposed methodology consists in predicting for each flight the trajectory to be followed for different weather forecasts and then calculating the total entry counts for each different scenario. The authors apply this methodology to the LECMSAU ATC sector and found that "for this particular example the uncertainty in the entry times increases as the distance travelled by the aircraft to the entry point increases".

In relation with the uncertainty in the trajectories, the SESAR project COPTRA developed a methodology for the probabilistic estimation of demand based on the combination of probabilistic trajectories. Taking into account the trajectory uncertainty calculated by applying polynomial chaos theory, the result of the combination is a probabilistic distribution of the entry counts in each sector [15].

4 ATFCM use cases selected for the project experiments and their BIM

To demonstrate the value and feasibility of the AICHAIN Technological Solution in ATM, the WP3 will explore two operational use-cases of interest for ATFCM/DCB, where the privacy-preserving private data exploitation enabled by AICHAIN can potentially bring significant operational value.

These two use-cases (UCs) are:

- UC1: ETOT prediction
- UC2: Pre-tactical traffic demand prediction

In this chapter the two use cases will be described in more detail and their preliminary benefit impact mechanism (BIM) analysed. Both use cases have been partially developed by ECTL and NOMMON, and in this project they will be enhanced with private data and tested in the federated ML solution through the AICHAIN prototype.

Section 4.1 presents UC1 and its BIM, and section 4.2 presents UC2 and its BIM. Note that the description of the selected use cases in this chapter is made from an operational point of view and the rationale from the ATM performance point of view. Following chapters will describe the ML models and datasets for these two use cases.

4.1 Operational description and rationale of UC1

The aim of the UC1 is to contribute to the improvement of the demand profile, through a better prediction of the take-off time. This will in turn contribute to [28]:

- Reduce the capacity buffers at sectors (less buffers will lead to higher declared capacities)
- Reduce the number of regulations (better precision and confidence in the assessment of potential sector overloads, in addition to the increased declared capacities)
- Move the demand and capacity management towards complexity management (ATCFM/DCB measures could be applied at the level of 4D trajectories instead of at the level of flows).

Today the Estimated Take-Off Time (ETOT) of each individual flight is obtained from the Enhanced Tactical Flow Management System (ETFMS) Flight Data (EFD), which is regularly updated from the submission of the Initial Flight Plan (IFP) to the Actual Take-Off Time (ATOT). The ETOT reported in the EFD, however, is not accurate enough to provide acceptable trajectory predictability. There exist several causes of the discrepancy between ETOT and ATOT, including congestion and bad weather at the airport, reactionary delays and Air Traffic Flow Management (ATFM) regulations.

A study published by EUROCONTROL [26], [27] demonstrated quantitatively that improved take-off predictability reduced the potential for sector over-delivery which in turn, could result in the reduction of en route sector buffers without compromising levels of safety. The study concluded that, following a wider implementation of Airport CDM, the benefits would be:

- It could be possible to **increase sector capacity** within the core area by up to 4% which equates to between 1-2 aircraft per sector
- A room for improvement for an en-route delays of between 33%-50%.
- Some sectors which are expected to be saturated are not actually saturated. Therefore if the declared capacities are maintained then some regulations may not be required.

A more recent study [29] updated and refined the previous analysis and reaffirmed that increasing ETOT predictability may lead to increase the sector capacities (although also pointed that the benefits could be less significant than the ones suggested in the previous study).

Figure 4-1 illustrates the relationship between higher predictability and the possibility to reduce the buffers in the declared/published sector capacities while maintaining the same levels of safety.

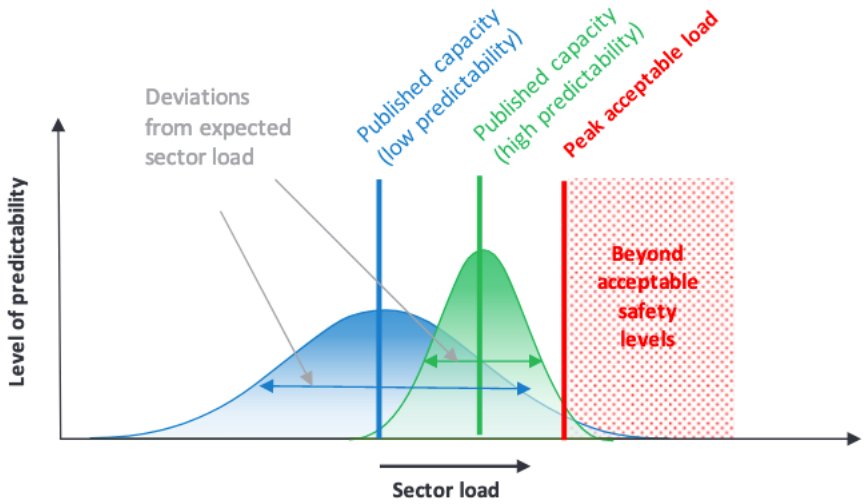


Figure 4-1 Higher predictability can lead to more sector capacity (source: the Airspace Architecture Study)

Regarding the complexity management possibilities, it is worth noting the Maastricht Upper Area Control Centre (MUAC) has developed the Traffic Predictions Improvements (TPI) project with the purpose to optimise the usage of ATCOs time and to reduce uncertainty of capacity predictions. The project aimed at improving predictability of traffic, allowing MUAC to make a more accurate view of the incoming traffic and consequently apply a better management of capacity and a better anticipation of the ATFCM and ATC measures on the demand. To achieve the goal MUAC uses innovative artificial intelligence (AI) methods to address the following sub-problems: route prediction, 4D trajectory prediction and sector sequence prediction.

The variability of entry times comes from factors external to MUAC, such as clearances by upstream ATC, general behaviour of aircraft outside MUAC airspace, or the take-off time uncertainty at the departure airport.

A simplified version of the ETOT prediction that was deployed in the TPI will be used in this project, in order to have a baseline for benchmarking federated ML versus non-federated ML with a relative realistic operational benchmark.

4.1.1 Preliminary BIM of the Use Case 1

4.1.1.1 Focus stakeholder: NM and ANSPs (ATFCM actors)

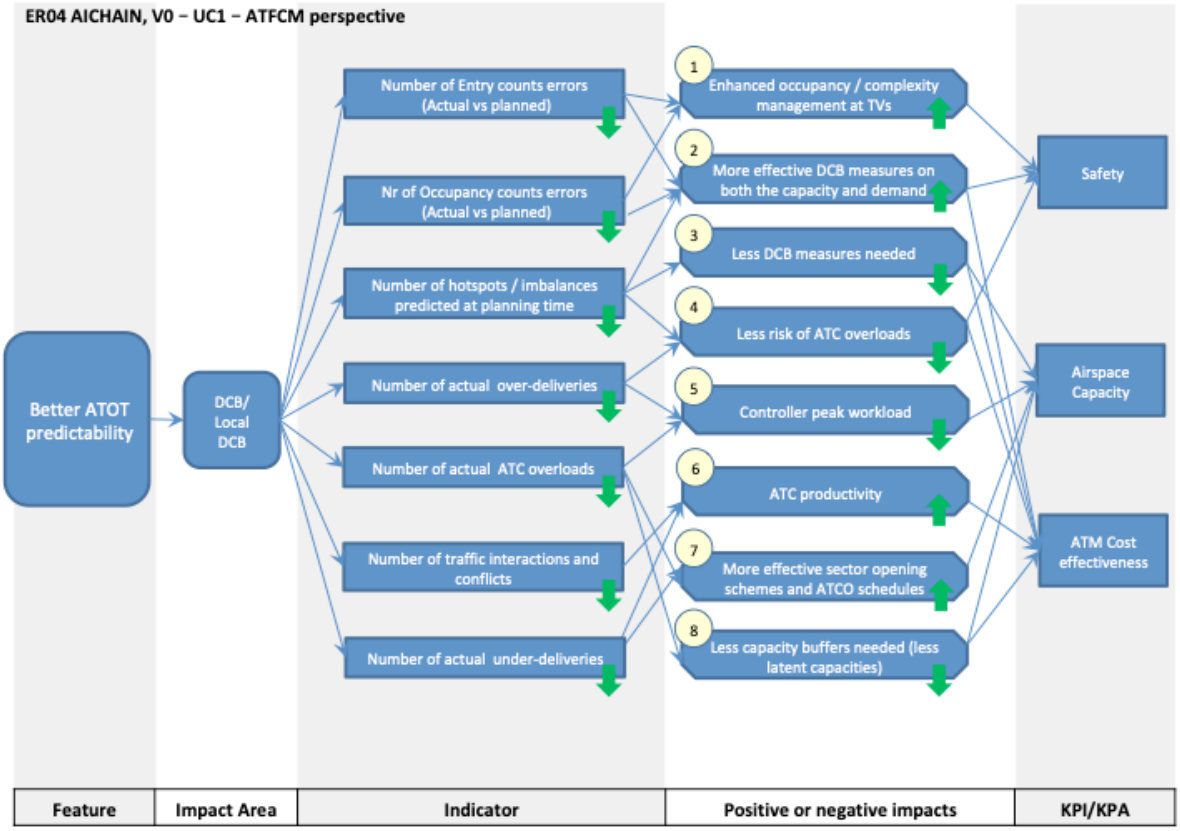


Figure 4-2 Benefit Impact Mechanism for Use Case 1 – ATFCM perspective

Contextual description of the BIM:

1. If the accuracy of the ETOT prediction is under a certain threshold (e.g. error below 5 min), then the more accurate entry counts and occupancy counts could enable advanced ATFCM management of complexity.

2. Higher ETOT accuracy means more robustness to uncertainties, which in turn means that the ATFCM decisions can be more effective and efficient, e.g. the sector configurations and opening scheme, the number of controllers allocated, the STAM scenarios and regulations, etc.

3. If the traffic counts and the ATFCM measures are more precise and effective, the natural consequence should be that the FMPs will have more control and will achieve more stability on the network, which should reduce the number of non-effective measures.

4. More effective ATFCM measures and more accuracy to detect potential ATC overloads can reduce the risk of ATC overloads.

5. Higher accuracy of counts and higher control on them should lead to less potentially dangerous workload peaks, while period of sustainable workload levels should increase.

6. With more effective DCB measures and less uncertainty in the traffic counts, the capacity buffers could be reduced, thus allowing more flights to be managed per hour with the same ATC resources (productivity increase).

7. Less uncertainty in the entry and occupancy counts could lead to sector opening schemes that are more robust, effective and cost-efficient.

8. With less uncertainty in the traffic counts, the capacity buffers could be reduced, thus reducing the latent capacities in the network.

4.1.1.2 Focus stakeholder: Airspace Users

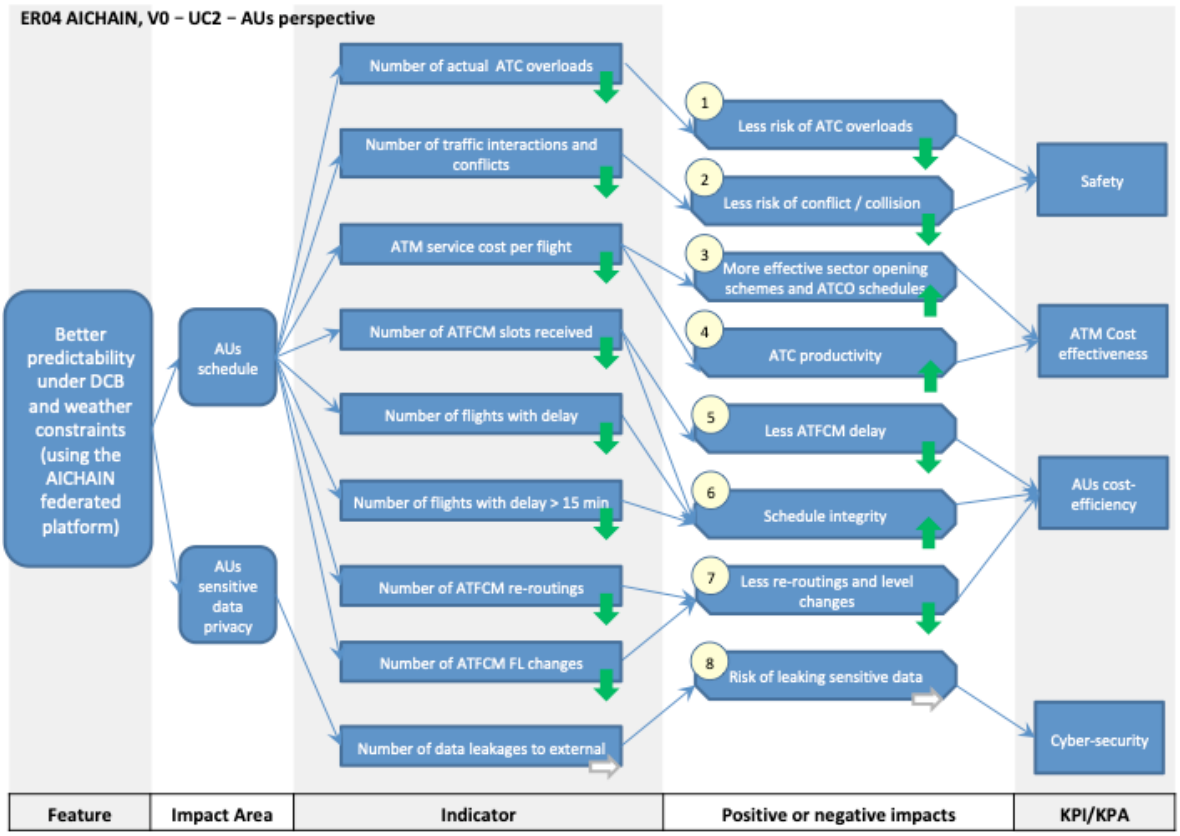


Figure 4-3 Benefit Impact Mechanism for Use Case 1 – AUs perspective

Contextual description of the BIM:

1. Higher accuracy in the traffic counts and more effective DCB measures should reduce the risk of ATC overload, which would increase the safety of the AUs' flights.

2. More effective density and complexity management measures, as well as the potential introduction of strategic de-confliction of 4D trajectories, should lead to less likeliness of airborne losses of separation and thus less risk of airborne collisions.

3. More effective sector opening schemes would have a positive impact on AUs operations in terms of less unnecessary constraints and less indirect costs per flight (i.e. higher ATM cost-effectiveness).

4. The increase of ATC capacity due to the reduction of latent capacities together with more effective plans of how to allocate the capacity available would lead to an increase of ATC productivity and consequently to a reduction of the indirect costs per flight (i.e. higher ATM cost-effectiveness).

5. A better dimensioned ATM system with higher capacity would lead to less ATFCM delay in the network (assuming the same level of traffic) which can be translated in less direct costs per flight (i.e. higher operational efficiency for the AUs).

6. Less ATFCM constraints in the network and less average delay per flight would reduce the number of cases in which the AUs schedules are disrupted, and therefore it would reduce the number of situation that can provoke very high costs to AUs during the day of operations (i.e. higher operational efficiency for the AUs).

7. The STAMs is often a good alternative to ATFCM delay because they can mitigate the impact of ATFCM delay on the AUs schedules. However, higher predictability in the traffic counts leading to more capacity and more effective ATCM measures could reduce the need for applying STAMs (e.g. re-routing or FL capping, among others), which would reduce the operational costs of AUs.

8. Assuming that the AICHAIN targeted architecture is feasible to enable privacy-preserving federated machine learning, the cyber-security risks of exploiting AUs private data in a federated setting is considered neutral since privacy and control of sensitive data is preserved.

4.2 Operational description and rationale of UC2

UC2 aims to improve the trajectory prediction with respect to last FPL sent during pre-tactical phase. Apart from predicting the expected flown route under normal conditions, the inputs of the tool shall be configurable to enable the introduction of hypothetical scenarios that can contain severe weather conditions or DCB constrains. The model will learn from AU historical data to take actions according to expected scenarios where weather or DCB constrains may apply. This Use Case is based on **PRETA solution developed by Nommon** for pre-tactical trajectory prediction, which has the following capabilities:

- Considers new data sources for enhanced features compared to the current tool (PREDICT)
- Better prediction accuracy rate for 3D routes, outperforming current solution up to 6.2%
- Enhanced seamless Demand Capacity Balance (DCB) functionalities.

- Potential integration with take-off time prediction for full 4D trajectory prediction.

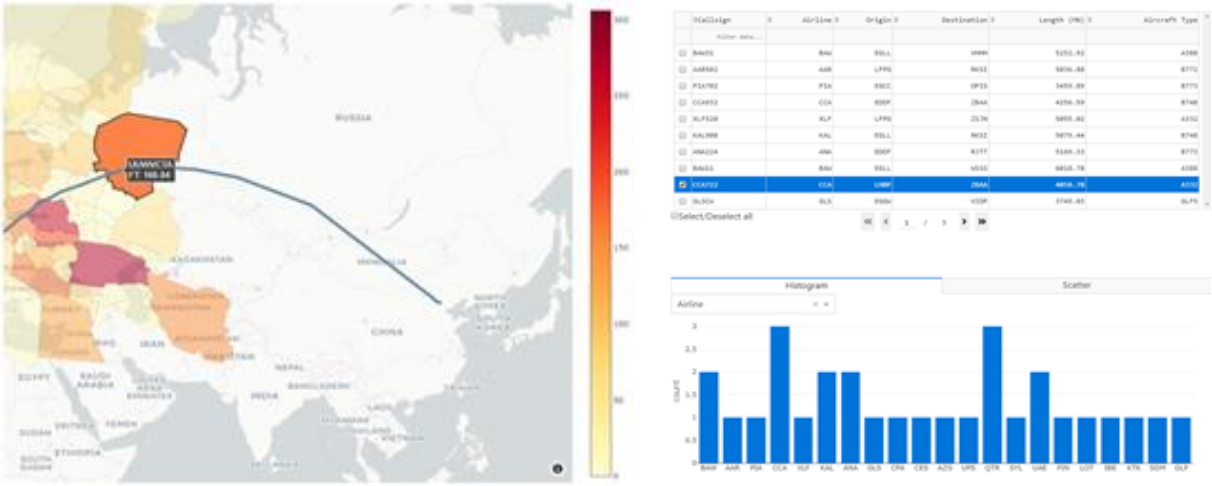


Figure 4-4 Main view of Nommon´s PRETA solution interface

The reasons that motivated the development of this Use Case are presented below:

- Actual demand is difficult to predict (e.g., due to weather) as a result of the high associated uncertainty.
- AUs constantly react to adapt during the day of operations, so ATFM cannot fully trust on the FPLs.
- The availability of better predictions of AU reactions could facilitate resource planning and allocation.

The proposed solution to overcome the current situation consists on the development of a Federated Learning (FL) model that, based on a set of stakeholder variables (the historic trajectories, the weather phenomena, the DCB constrains and the AU´s federated data) is able to accurately forecast the trajectory each flight is going to follow during its operations.

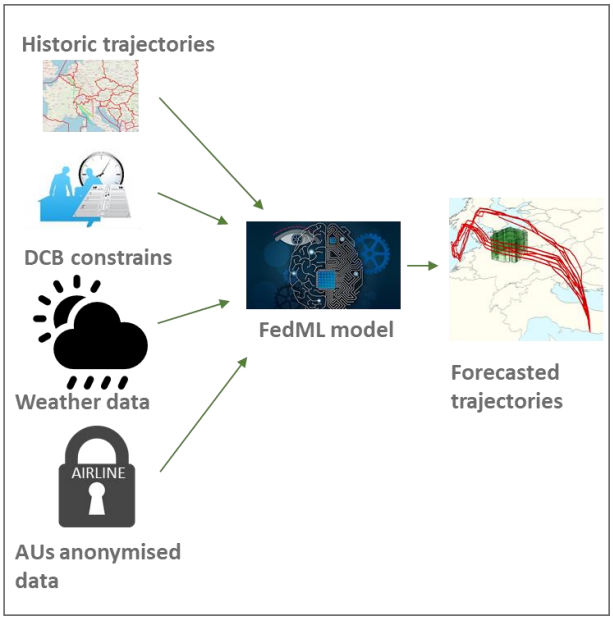


Figure 4-5 UC2 FedML model conceptual design

It is expected that the proposed machine learning model will be able to increase the predictability of AU´s behaviour, facilitating DCB planning and hence achieving a better resource allocation by making an optimal use of ATFM regulations.

4.2.1 Preliminary BIM of the Use Case 2

4.2.1.1 Focus stakeholder: NM and ANSPs (ATFCM actors)

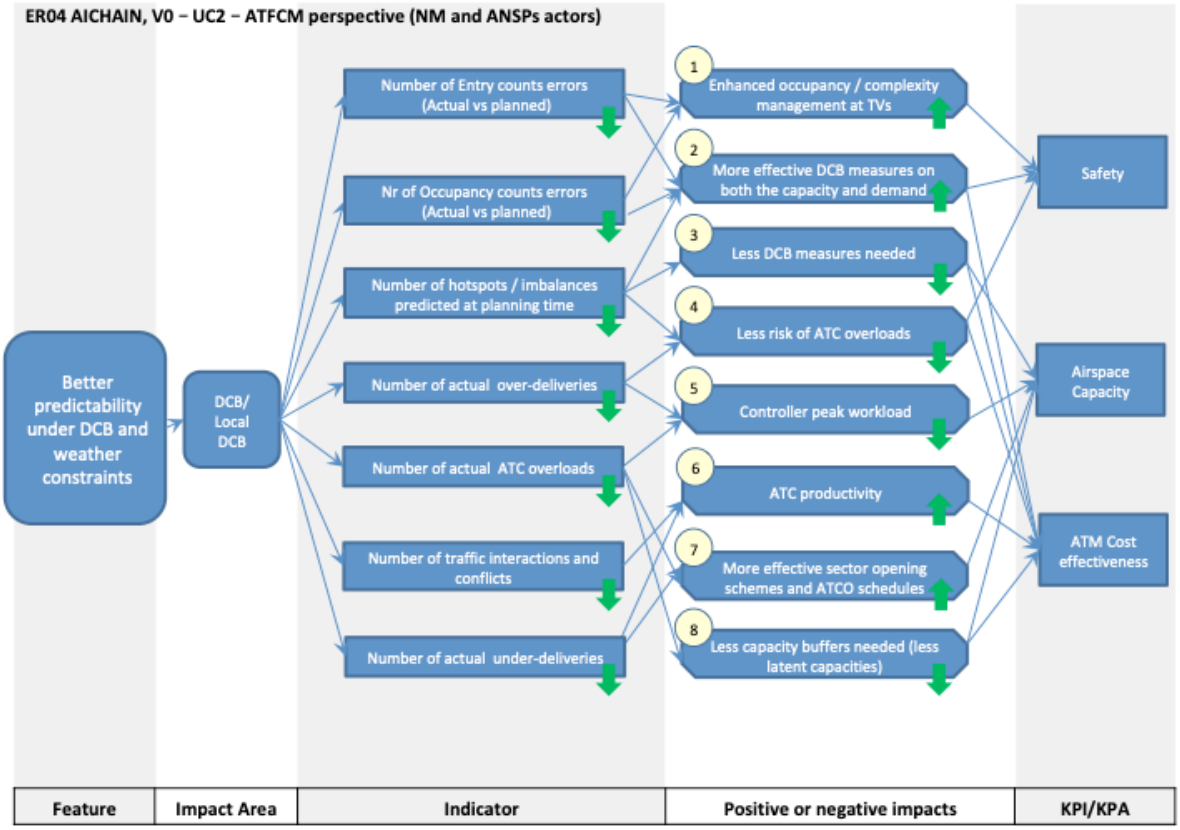


Figure 4-6 Benefit Impact Mechanism for Use Case 2 – ATFCM perspective

Contextual description of the BIM:

1. The increment of the trajectory prediction accuracy translates into more accurate entry counts and occupancy counts that could enable advanced ATFCM management of complexity.

2. Higher trajectory prediction accuracy under severe weather or DCM measures means more robustness to uncertainties, which in turn means that the ATFCM decisions can be more effective and efficient, e.g. the expected effect of a STAM could be predicted so network effects could be avoided.

3. If the traffic counts and the ATFCM measures are more precise and effective, the natural consequence should be that the FMPs will have more control and will achieve more stability on the network, which should reduce the number of non-effective measures.

4. More effective ATFCM measures and more accuracy to detect potential ATC overloads can reduce the risk of ATC overloads.

5. Higher accuracy of counts and higher control on them should lead to less potentially dangerous workload peaks (even during severe weather or capacity imbalances situations), while period of sustainable workload levels should increase.

6. With more effective DCB measures and less uncertainty in the traffic counts, the capacity buffers could be reduced, thus allowing more flights to be managed per hour with the same ATC resources (productivity increase).

7. Less uncertainty in the entry and occupancy counts could lead to sector opening schemes that are more robust, effective and cost-efficient.

8. With less uncertainty in the traffic counts, the capacity buffers could be reduced, thus reducing the latent capacities in the network.

4.2.1.2 Focus stakeholder: Airspace Users

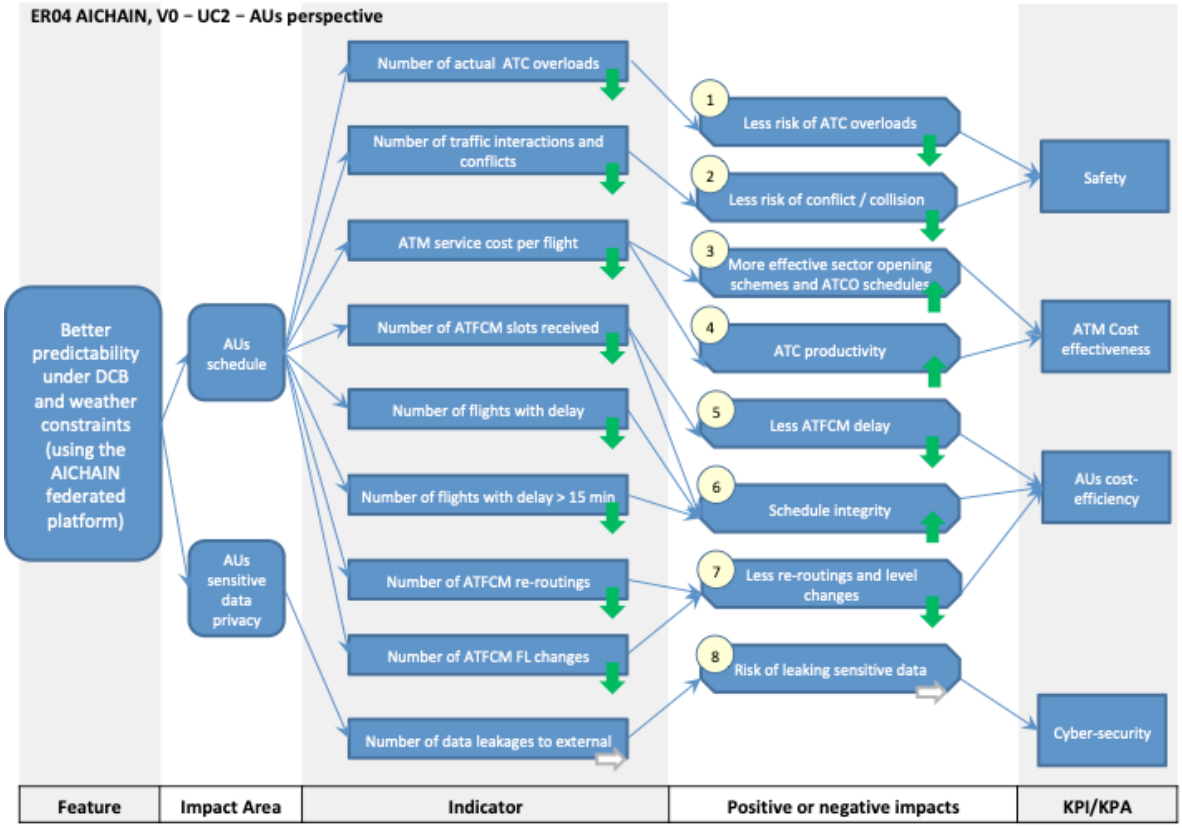


Figure 4-7 Benefit Impact Mechanism for Use Case 2 – AUs perspective

Contextual description of the BIM:

1. Higher accuracy in the traffic counts and more effective DCB measures should reduce the risk of ATC overload, which would increase the safety of the AUs' flights.

2. More effective density and complexity management measures, as well as the potential introduction of strategic de-confliction of 4D trajectories, should lead to less likeliness of airborne losses of separation and thus less risk of airborne incidents.

3. More effective sector opening schemes would have a positive impact on AUs operations in terms of less unnecessary constraints and less indirect costs per flight (i.e. higher ATM cost-effectiveness).

4. The increase of ATC capacity due to the reduction of latent capacities together with more effective plans of how to allocate the capacity available would lead to an increase of ATC productivity and consequently to a reduction of the indirect costs per flight (i.e. higher ATM cost-effectiveness).

5. A better dimensioned ATM system with higher capacity would lead to less ATFCM delay in the network (assuming the same level of traffic) which can be translated in less direct costs per flight (i.e. higher operational efficiency for the AUs).

6. Less and more efficient ATFCM constraints in the network and less average delay per flight would reduce the number of cases in which the AUs schedules are disrupted, and therefore it would reduce the number of situations that can provoke very high costs to AUs during the day of operations (i.e. higher operational efficiency for the AUs).

7. The STAMs is often a good alternative to ATFCM delay because they can mitigate the impact of ATFCM delay on the AUs schedules. However, higher predictability in the traffic counts leading to more capacity and more effective ATCM measures could reduce the need for applying STAMs (e.g. re-routing or FL capping, among others), which would reduce the operational costs of AUs but also the environmental impact by reducing the number of inefficient re-routings.

8. Assuming that the AICHAIN targeted architecture is feasible to enable privacy-preserving federated machine learning, the cyber-security risks of exploiting AUs private data in a federated manner is considered neutral since privacy and control of sensitive data is preserved.

5 Models and datasets description for UC1

5.1 ML canvas description

Table 5-1 Canvas description of UC1

ML Canvas:		Description of UC1
Value Proposition *(Goals, What, Why, Who)*		*The goal of the UC1 model is to improve the accuracy of the take-off time prediction, i.e., the ETOT (expected take-off time), with respect the one calculated by the ETFMS system. Trained against (ATOT - ETOT) target, this model would provide an improved ETOT value based on the initial ETOT (and other features as input). This prediction will then contribute to the improvement of the Demand Profile. A more accurate ETOT can help the ANSPs and the NM to better assess the traffic demand at sectors. For some short-term tactical planning timeframes (e.g. less than 1 hour) more accurate ETOTs may facilitate the management of traffic complexity and even trajectory interactions (strategic conflict management). This should result in less unnecessary constraints on the traffic demand and in a better planning of the capacity resources, thus potentially leading to a reduction of the latent capacities (less capacity buffers).*
LEARN	Data Sources	The data sources for the model will come from the ETFMS (internal system of NM) enhanced with the federated datasets in the airspace users' facilities.
	Features (Engineering)	The complete set of features required by the model is detailed in the following sections for both the internal NM data and the external airspace users data.
	Collecting Data	New data will be collected through the NM data provision services and the airspace users federated datasets. The ETFMS will provide both the baseline ETOT (the one to be enhanced) and the ATOT (actual take-off time), to train the model and monitor the performance.
	Building Models	In the current operational models used in MUAC to enhance the ETOT estimated by the ETFMS is re-trained every 2 months approximately. In this project is out of the scope to provide any recommendation with respect how frequent the model should be re-trained. Therefore one single training should be assumed.
PREDICT	Prediction Task	The model will take as inputs the baseline ETOT calculated by the ETFMS, and will use the rest of the data features to estimate the error of the ETOT with respect the ATOT, i.e. (ATOT – ETOT) target. The error calculated by the model added to the ETOT should result in a new more ETOT (closer to the ATOT).

	Offline Evaluation	The evaluation will be conducted by splitting the dataset in training set, dev set and test set (60%, 20%, 20% approx.). The error will be calculated using the mean average error (MAE) and the root mean square error (RMSE), as the two metrics have different sensitivity to outliers The performance will be analysed as a function of the look ahead time in addition to the one in aggregated form.
	Decisions	Decisions made by NM or the ANPSs will use the enhanced ETOT predictions to generate operational value could be related to: a) the traffic density and/or complexity management; b) trajectory interactions management (strategic deconfliction); c) capacity planning (e.g. capacity buffers reduction); or d) capacity management and sector opening scheme. Airlines and airport operations could benefit as well from more accurate ETOT.
	Making Predictions	There are two possible setups in which NM can make the take-off time predictions in a federated platform: a) with the data that is visible to NM only; or b) with both the NM data and the federated private datasets of the AUs. In this project the second setup will be assumed.
Live Evaluation and Monitoring		In an operational version of the model, the real-time monitoring and evaluation could be limited to the assessment of the prediction error as soon as the ATOT becomes available, i.e., ETOT vs ATOT error. However, in the context of this project it is out of the scope to reproduce and assess any real time operational environment. Thus, no live assessment will be conducted.

5.2 Data sets of the non-federated model (dataset in the NM side)

The next three tables present the features that have been used in the ETOT prediction model grouped in three classes, i.e. Flight, weather and regulated traffic volumes crosses by a regulated flight.

The type of feature can be *raw* or *made.* 'raw' means that the variable was present in the data source and 'made' means that the variable has been engineered from the raw data.

Some parameters are *dynamic,* which means that the information associated changes with time, and thus several samples (with different time-stamps) are needed to characterise a flight, a regulation or a weather object.

The features with the "_LEG" suffix are computed by using information from the previous flight leg operated by the same aircraft (i.e., tail number). For instance, if a given aircraft is performing the hypothetical rotations A → B → C → A during the day (where A, B and C are airports), the value of

the FLIGHT_DURATION_LEG when predicting the take-off time of flight B corresponds to the duration of flight A. Note that, since night-stops may last several hours, a threshold of one day (24 hours) between the EOBT of the flight for which the take-off time is predicted and the ATA of previous flight leg operated by the same aircraft has been used to link two flights. In the event that this condition is not met, then the features ending by the suffix "_LEG" are *missing* in the corresponding sample. Fortunately, gradient boosted decision trees (GBDT), the type of model used in this experiment, handles missing values very efficiently, by simply ignoring the features that are not present when performing the prediction.

Furthermore, note that airspace users may change the aircraft assigned to a given flight at any moment by performing what is commonly known as an aircraft *swap.* This strategy is often used to avoid the propagation of delay to critical flights (e.g., those with high risk of infringing the night curfew at the destination airport or with many connecting passengers). For this reason, the ARCTYP feature shown in Table 2 is considered dynamic. In turn, the features with the "_LEG" suffix are also dynamic, and are updated whenever the aircraft assigned to the current flight changes taking into account the previous flight operated by the new aircraft.

Finally, in most of Europe, standard instrument departures (SIDs) are named after the final waypoint of the procedure, followed by a version number that is incremented by one each time the procedure is modified, and a single letter that designates the associated runway. For instance, LOPIK1F refers to a SID of Amsterdam-Schipol airport (EHAM) that connects runway 04 with the LOPIK waypoint. Accordingly, the last letter of the SID (F in the example of the previous sentence) can be used as a proxy for the expected take-off runway. This simplification has been used because the take-off runway is not directly present in the EFD messages.

Table 5-2 Flight features

Feature	Type	Dynamicity	Description
EVENT	raw	Dynamic	This field provides the source of the EFD. This can be an incoming message, a user command or a system event, which is known to the ETFMS users.
EVENTCLASS	raw	Dynamic	This field can contain the following values: "MSG" - the EFD source is an incoming or outgoing message "MAN" - the EFD is triggered by an FMD user command "SYS" - the EFD is automatically generated by a time trigger event "REG" - the EFD is automatically generated by a slot re-calculation event
FLTSTATE	raw	Dynamic	The status of the flight, which can be: "FI" - Filed. Basic status. "FS" - Filed_Slot_allocated The flight is regulated, but the slot has not yet been published via a SAM. "SI" - Slot_Issued. The flight is regulated and the slot has been published via a SAM.

			"TA" - Tact_Activated. ETFMS assumes that the flight is airborne, but it has not yet received a confirmation from ATC (yet). "AA" - Atc_Activated.
ADEP	raw	Static	Aerodrome of departure of the flight.
ADES	raw	Static	Aerodrome of destination of the flight.
ARCTYP	raw	Dynamic	Aircraft Type (ADEXP)
IRULES	raw	Static	Initial flight rules, initial flight type and initial IFPS processing indication of the flight.
RDYSTATE	raw	Dynamic	Ready status of the flight. It can be: "D" - ready to Depart (REA message received). “N“ - not ready to Depart (no REA received yet). “I” - ready for Improvement “S” - SIP wanted
TAXITIME	raw	Dynamic	The taxitime-field contains the most recently known taxitime value by ETFMS.
AOARCID	raw	Static	It contains the 3 letter ICAO designator of the AO which is derived from the ARCID of the flight plan
FLTTYP	raw	Static	Flight Type. S: Scheduled N: Non-scheduled G: General aviation M: Military X: Other types
DEPAPTYPE	raw	Static	Type of aerodrome of departure, which can be: ADVANCEDATCTWR: to indicate an airport that has been classified in NMOC as an advanced ATC TWR (which can send e.g. ATC-DPI messages). CDM: to indicate an airport that has been classified in NMOC as a CDM airport (and which can send e.g. send all types of DPI messages. Note that this APTYPE is also used for AOP-NOP airports
CDMSTATUS	raw	Dynamic	cdmstatus values as available in ETFMS (not present for non-CDM airports). Values can be: DPIEXPECTED: Default value. DPI messages are expected. PREDICTED: The P-DPI message has been received. ESTIMATED: The E-DPI message has been received.

			TARGETED: The T-DPI-t message has been received. PRESEQUENCED: The T-DPI-s message has been received. ACTUALOFFBLOCK: The ATC-DPI message has been received.
AOOPR	raw	Static	It contains the 3 letter ICAO designator of the AO which is derived from the OPR/ field in the field18 of the flight plan.
ATFMDELAY	raw	Dynamic	The ATFM delay allocated by the ETFMS system to that flight. This value is calculated at the time the EFD was transmitted. It corresponds to CTOT that has been issued by the ATFM messages SAM/SRM
IFPSDISCREPANCY_REG	raw	Dynamic	The Aircraft Registration received from the departure airport is different from the Aircraft Registration in the ICAO flight plan or is missing in the ICAO Flight plan.
IFPSDISCREPANCY_ARCTYP	raw	Dynamic	The Aircraft Type received from the departure airport is different from the Aircraft Type in the ICAO flight plan.
IFPSDISCREPANCY_OBT	raw	Dynamic	The off-block time of the ICAO flight plan is more than 15min different from the off-block time derived from E-DPI or T-TDPI-t messages.
DEPSTATUS	raw	Dynamic	Special status of the flight at the departure airport. Can be: DEICING: The Aircraft is being de-iced or will be de-iced.
ADESOLD	raw	Dynamic	The adesold-field concerns a flight that has been diverted to a new aerodrome of destination and contains the aerodrome of destination that was filed in the flight plan.
RWY	made	Dynamic	Last letter of the Standard Instrumental Departure (SID) procedure. Used as a 'proxy' for the take-off runway.
FLIGHT_DURATION	made	Dynamic	Scheduled duration of the flight.
EOBT_IFP_TO_EOBT	made	Dynamic	Difference between the current Estimated Off-Block Time (EOBT) and the EOBT according to the Initial Flight Plan (IFP).
ADEPETO_IFP_TO_ADEPETO	made	Dynamic	Difference between the current Estimated Take-Off Time (ETOT) and the ETOT according to the IFP.
TIMESTAMP_IFP_TO_TIMESTAMP	made	Dynamic	Time from the reception of the IFP at the ETFMS.
TIMESTAMP_TO_ADEPETO	made	Dynamic	Time to ETOT.
TIMESTAMP_TO_EOBT	made	Dynamic	Time to EOBT.
TIMESTAMP_TO_TSAT	made	Dynamic	Time to Target Start-up Approval Time (TSAT).
TIMESTAMP_TO_TOBT	made	Dynamic	Time to Target Off-Block Time (TOBT).
TIME_FROM_REG_CHANGE	made	Dynamic	Time from allocation or change of aircraft registration number (i.e., airframe) to the flight.

TURNAROUND_LEG	made	Dynamic	Available turn-around time. This feature is computed as the difference between two terms (A-B): A: The Controlled Off-Block Time (COBT) / EOBT, for regulated and non-regulated flights, respectively. B: The Actual Time of Arrival (ATA) / Controlled Time of Arrival (CTA) / Estimated Time of Arrival (ETA) of the previous flight leg, for terminated (TE), regulated and non-regulated flights, respectively. Note: this value is negative if the ATA / CTA / ETA of the previous flight is later than the COBT / EOBT of the current leg.
FLIGHT_DURATION_LEG	made	Dynamic	Duration of the previous flight leg
TIMESTAMP_TO_TA_LEG	made	Dynamic	Time to the ATA / CTA / ETA of the previous flight leg. Note: this value is negative when the previous is airborne (AA) or TE
ADEPETO_IFP_TO_ADEPETO_LEG	made	Dynamic	Delay of the previous flight leg. Computed as the difference between Actual Take-off Time (ATOT) or most recent ETOT of the previous leg and the ETOT of its corresponding IFP.
EOBT_IFP_TO_EOBT_LEG	made	Dynamic	Same as EOBT_IFP_TO_EOBT but for the previous flight leg.
TIMESTAMP_LEG_TO_TIMESTAMP	made	Dynamic	Time from last EFD message sent by the previous flight leg.
EVENT_LEG	made	Dynamic	Last source of the EFD sent by the previous flight leg (see EVENT feature)
AOOPR_LEG	made	Dynamic	Aircraft operator of the previous flight leg derived from the OPR/ field in the field18 of the flight plan.
FLTSTATE_LEG	made	Dynamic	Last flight status of the previous flight leg (see FLTSTATE feature).
ADEP_LEG	made	Dynamic	Aerodrome of departure of the previous flight leg.
AOARCID_LEG	made	Dynamic	Aircraft operator of the previous flight leg derived from the ARCID of the flight plan.
FLTTYP_LEG	made	Dynamic	Flight Type of the previous flight leg (see FLTTYP feature)
HOUR	made	Dynamic	Hour of the day
MONTH	made	Dynamic	Month of the year
DAY	made	Dynamic	Day of the week

Table 5-3 Weather features

Feature	Type	Dynami-city	Description
precipitation_intensity	raw	Dynamic	Intensity of the precipitation: –: Light +: Heavy VC: In the vicinity
precipitation_description	raw	Dynamic	Description of the precipitation: MI: Shallow BC: Patches PR: Partial DR: Low drifting BL: Blowing SH: Shower(s) TS: Thunderstorm FZ: Freezing
precipitation	raw	Dynamic	Type of precipitation: DZ: Drizzle RA: Rain SN: Snow SG: Snow grains PL: Ice pellets GR: Hail GS: Small hail and/ or snow pellets UP: Unknown
obscuration_intensity	raw	Dynamic	Intensity of the obscuration (same categories as precipitation intensity).
obscuration_description	raw	Dynamic	Description of the obscuration (same categories as precipitation description).
obscuration	raw	Dynamic	Type of obscuration: BR: Mist FG: Fog FU: Smoke VA: Volcanic ash DU: Widespread dust

			SA: Sand HZ: Haze
vv	raw	Dynamic	Vertical visibility (in ft)
cover	raw	Dynamic	Cloud cover: FEW: few SCT: scattered BKN: broken OVC: overcast
height	raw	Dynamic	Height of the lowest layer of clouds (in ft)
cloud	raw	Dynamic	Type of cloud CB: Cumulonimbus TCU: Cumulus congestus of great vertical extent.
visibility	raw	Dynamic	Visibility (in SM)
temperature	raw	Dynamic	Temperature (in K)
pressure	raw	Dynamic	Pressure (in hPa)
dewpt	raw	Dynamic	Dew point (in K)
windSpeed	raw	Dynamic	Wind speed (in kt)
windDir	raw	Dynamic	Wind direction (in deg)
windGust	raw	Dynamic	Wind gust (in kt)
snowDepth	raw	Dynamic	Snow depth (in inches)

Table 5-4 Regulated traffic volume features

Feature	Type	Dynami-city	Description
TV_0	raw	Dynamic	Traffic volume of the most penalising regulation
CAUSE_0	raw	Dynamic	Cause of the most penalising regulation
TV_1	raw	Dynamic	Traffic volume of the 2nd most penalising regulation
CAUSE_1	raw	Dynamic	Cause of the 2nd most penalising regulation
TV_2	raw	Dynamic	Traffic volume of the 3rd most penalising regulation
CAUSE_2	raw	Dynamic	Cause of the 3rd most penalising regulation
TV_3	raw	Dynamic	Traffic volume of the 4th most penalising regulation
CAUSE_3	raw	Dynamic	Cause of the 4th most penalising regulation
TV_4	raw	Dynamic	Traffic volume of the 5th most penalising regulation
CAUSE_4	raw	Dynamic	Cause of the 5th most penalising regulation

5.3 Data sets of the federated model (dataset in the AU side)

The following tables show the features gathered from the AU's side that contain private and sensitive information that normally is not available to the NM. The features can be grouped in the following groups: a) Estimated Times (estimated by the AU); b) Passengers, cockpit and crew; c) Aircraft; and d) Airport.

Table 5-5 Estimated times (estimated by the AU)

Feature	Type of feature	Type of availability	Comments
EOBT	Datetime	Dynamic	
ETOT	Datetime	Dynamic	
EXIT (taxi-in)	Datetime	Dynamic	
EXOT (taxi-out)	Datetime	Dynamic	
Expected turn-around time	Numerical	Static	

Table 5-6 Passengers, cockpit and crew;

Feature	Type of feature	Type of availability	Comments
Passenger boarding status (overall flight)	Categorical	Dynamic	Values: Open, Closed, FirstPassengerBoarded
Luggage boarding status (overall flight)			WIP
Number of passengers	Numerical	Dynamic	Booked and boarded
Passenger connections	Numerical	Dynamic	Numerical with different logical categories

Cockpit / crew rotation	Numerical	Dynamic	Numerical with different logical categories
Change of cockpit / crew rotation	Binary	Dynamic	ZRH and GVA only

Table 5-7 Aircraft

Feature	Type of feature	Type of availability	Comments
Aircraft type	Categorical	Dynamic	Example: A330
Change of aircraft type	Binary	Dynamic	
Aircraft capacity	Numerical	Static	Example: 3/150
Aircraft maintenance plan			WIP, not available for the moment
Flight priority within AU			WIP, not available for the moment

Table 5-8 Airport

Feature	Type of feature	Type of availability	Comments
Gate assigned	Binary	Dynamic	ZRH and GVA only
Gate area	Categorical	Dynamic	ZRH and GVA only. Example: A15
Change in gate assigned	Binary	Dynamic	ZRH and GVA only
Occupancy of parking stands	Numerical	Dynamic	ZRH and GVA only. Fraction of occupied parking stands at departure airport
Runway configuration	Categorical	Dynamic	ZRH only. Example: North Concept

6 Models and datasets description for UC2

6.1 ML Canvas description

Table 6-1 Canvas description of UC2

ML Canvas:		Description of UC1
Value Proposition *(Goals, What, Why, Who)*		The model proposed in UC2 shall help the NM to improve its knowledge regarding actual flown 3D trajectories (2D route + flight level) with respect to the flight plan (FPL) and given the status of all routes for the current state of DCB constraints and weather conditions. This way, the NM will have a tool to predict the expected flown routes in normal conditions as well as simulating the route selection under hypothetical constraints on weather and DCB. As a result, the process of resource allocation could be optimised to adjust to the actual demand of the network which will result in a reduction of ATFM regulations. For that purpose, the proposed model will perform a prediction of the 3D trajectory that is actually selected using as features route details, DCB and weather constrains and the filled FPL itself.
LEARN	Data Sources	The data sources for the model are three: • **DDR-2** (EUROCONTROL´s demand data repository): provides the historic of trajectories and regulations. • **ECMWF** (European Centre for Medium-Range Weather Forecasts): provide the necessary weather products. • **Airspace user proprietary data** in federated datasets within each AU's facilities.
	Features (Engineering)	The complete set of features required by the model is detailed in the following sections for all the internal NM data, the external sources and the external airspace user's data.
	Collecting Data	Data collection is performed by each data source responsible partner and its update relies on them. In general, data updates will be directly incorporated to the system through the federated nodes at each party.
	Building Models	The general model will be composed of two classification model: A 2D route model, which will predict the 2D route flown among all available and a Flight Level model that will predict cruise the flight level (discrete).

		Even though, model update strategy is outside the scope of this project, it is expected for the model to be re-trained with monthly frequency, which is feasible as pre-processing and modelling steps take at most one day.
PREDICT	Prediction Task	The model will perform classification on two levels: 2D route from a list of available routes each given day and flight level out of the possible flight levels for each given OD pair. Each prediction will consist on a binary choice for each observed 2D route/flight level following the question: is this route selected? Features associated to each classification point will correspond to the variable observations constrained to the geographical area of the route/flight level.
	Offline Evaluation	The evaluation will be conducted by splitting the dataset in training set, test set and validation set (80%-20% approx. for train/test and a new AIRAC cycle for validation.). To validate the Machine Learning models internally (binary classification), accuracy and F1-score will be used. To validate final route prediction accuracy will be used.
	Decisions	The NM or ANPSs will use the enhanced 3D trajectory predictions to generate operational value with respect to: a) the traffic density and/or complexity management; b) trajectory interactions management (strategic deconfliction); c) capacity planning (e.g. capacity buffers reduction); or d) capacity management and sector opening scheme. Airlines and airport operations could be indirect beneficiaries of a more accurate trajectory forecast.
	Making Predictions	Data from all federated parties will be considered for predictions. Predictions will be requested by users on a flight-by-flight basis, even though batch predictions will be supported. While the prediction horizon will be short-term and predictions shall be computed in a short time period, real-time prediction is not expected.
Live Evaluation and Monitoring		In an operational version the model, the real-time monitoring and evaluation could be limited to the assessment of the prediction correctness and/or the impact of the error in other metrics (e.g., complexity indicators). However, in the context of this project it is out of the scope to reproduce and assess and real time operational environment. Thus, no live assessment will be conducted.

6.2 Datasets of the non-federated model

Taking into account the amount of effort and resources required to build a full-scale model (e.g., at ECAC level) and due to the limited resources available for this WP, the consortium the geographical scope of this problem has been limited to a single corridor relevant for SWISS as AU data provider (See Figure 6-1). Concisely, the selected flow consists on all the flights connecting Switzerland and the Greater London area. The flow meets all the requirements needed for the UC, as it is often affected by severe weather events, the number of flights is significant and SWISS has a relevant share of all the flights within

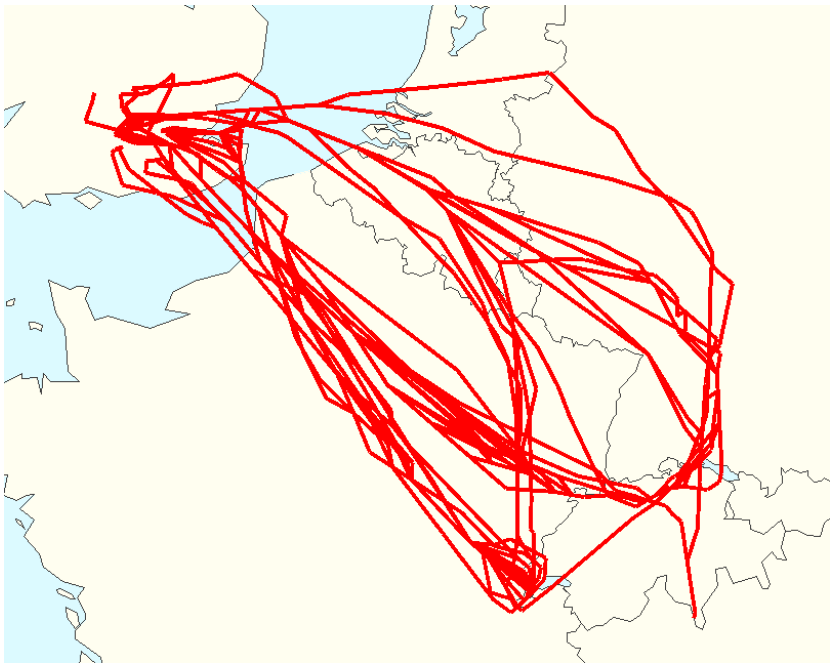


Figure 6-1 2D route representation for the selected geographical scope

The data sources to be considered in the UC2 has already been mentioned in the Canvas and are listed below:

- **Trajectory data:** mainly actual trajectory data although FPLs may be used, data will be obtained from DDR-2.
- **Regulation data:** en-route regulations also obtained from DDR-2 for the designated geographical area.
- **Weather data:** obtained from ECMW, provides historical data (GRIB2 files) containing metrics than can be used as weather proxies (k-index, total-totals). Data is available each 6 hours in a 0.5 geodesical degrees grid.

The feature selection for the UC2 is considered a model development task and will be undertaken in future tasks of WP3. Nevertheless, the UC2 model design presents a significant difference with UC1: no dynamic variables will be considered. The model will consider all variables as static, for example, the weather considered for the training will be the actual weather for each flight in the training dataset. While this issue could impact the performance of predictions, it is worth noting that the

intended use of this tool is as a What-if support tool that given some expectations (could be predictions or directly hypothetical scenarios) predicts how will each AU react.

6.3 Datasets of the federated model (dataset in the AU side)

The federated model will take as input the entire collection of non-federated features and a set of AU business-sensitive variables that will be included through federation. The following table summarises the AU variables that will be included in the model.

Table 6-2 UC2 federated variables

Feature	Description
TOW	Take-off weight for each flight
RUNWAY_CNFG	Runway configuration available for take-off and landing (only available for Zurich and Geneva)
CREW_PREVIOUS_TIME	Connection time from the crew´s previous flight (considering schedule time)
IN_PAX	Passengers of the flight fed form a connection flight
OUT_PAX	Passengers of the flight feeding a connection flight
AVE_MACH	Average Mach during cruise

7 AICHAIN Validation Plan

This chapter includes the Validation Plan of the AICHAIN project, and acts as an interim single-entry document to understand the full validation approach of the AICHAIN project (i.e. thus extending the scope to WP2 and WP4, and not only linked to WP3). At the end of the project this chapter will be included, after the due updates, in the last deliverable D1.6, which will act as the final single-entry document of the project with regards to the validation and experimentation approach.

7.1 General validation objectives

The underlying ATM concept researched in the AICHAIN project (i.e., *using privacy-preserving Federated Machine Learning over Blockchain technologies to enable the secure exploitation of private data and, as a consequence, enabling new operational improvements in the context of ATM*), can be declared **as pre-TRL1 maturity**, since up to the best of our knowledge no one explored this concept before in the ATM domain.

Since the AICHAIN solution is at pre-TRL1 maturity, the goal of the research and the validation plan will be **"to establish and quantify the need for the change"**, as stated in the SESAR reference validation methodology E-OCVM for V0 / pre-TRL1 concepts. In particular, the aim of the project is to develop a **portfolio of evidence** to let the SJU deciding whether the proposed AICHAIN technological solution should be carried to further research and development.

The portfolio of evidence will be build with as much as **quantitative and qualitative evidence** as possible, targeting the validation needs identified in the E-OCVM for V0 / pre-TRL1 concepts, and taking into consideration what is **realistically doable** given the maturity level of the concept as well as the **budget and time constraints** of the project.

7.2 General validation approach

Following the recommendations and research requirements of the call for proposals (topic 04 - DIM), the new DIM enabler will be explored, preliminarily, in its four dimensions, i.e., **operational, technological, economical and regulatory**. In the scope of this project we address only some of the *economic and regulatory* aspects in a common perspective to which we refer to as ***governance and incentives***, thus resulting in three research areas. Figure 7-1 shows the scope of the project, with three work packages, one per area of research, and how they are interrelated to each other from a high-level DIM system point of view.

Regarding the **technological dimension** of the AICHAIN solution, the project aims at demonstrating the technical feasibility of private data exploitation enabled by AICHAIN with advanced federated ML training techniques. In particular, the AICHAIN technical solution will contribute to **enable advanced ML models** that can be trained **without violating the privacy and data ownership** while guaranteeing the **data integrity and provenance**, among other requirements. **A small-scale prototype of the AICHAIN**

enabler will be a good demonstrator of these features and a valuable tool for preliminarily analysing the fulfilment of the requirements of the DIM enabler.

Regarding the **operational dimension,** it must be noted that **the ultimate purpose of AICHAIN** is to enable new SESAR operational improvements that can potentially increase the **predictability of traffic demand,** the **airspace capacity** and/or the **operational cost-efficiency**, among other ATM performance areas. **To demonstrate the potential operational value of the AICHAIN solution, two ATM case studies of interest for the Network Manager will be fully developed and researched in this project.**

The two use cases belong to the domain of **Advanced Demand and Capacity Balancing (A-DCB)** of the SESAR Industrial Research (IR) pipeline**,** and are today among the ones of highest interest and priority for the Network Manager (note however that the DIM enabler proposed could be useful in other contexts as well, e.g., advanced air traffic services, total airport management, or others). The identified A-DCB use cases are:

- **Prediction of expected take-off time (ETOT).** This use case is today being explored with NM data only. It may be interesting to research if FedML could improve the prediction performance of the current ML approaches by using privacy-protected data of Airspace Users.

- **Prediction of AUs' reaction (change of flight plans) to network and weather constraints.** This information could be a good input for FMPs and the NM before activating a network constraint (e.g., a regulation) to anticipate the potential reaction of AUs; in this way more efficient constraints could be designed, thus reaching more stability and predictability in the network and a better quality of service for AUs.

Regarding the **governance & incentive dimension**, it is relevant to know as soon as possible if the private data owners (e.g. the AUs) will be willing to participate in the collaborative training of the models, which is necessary condition for the success of the proposed solution. The scope of the research will be limited to the design –but not to the implementation– of a basic/fundamental incentive mechanism, which will be supported by qualitative evidence and formal analysis, and complemented with a well-justified proposition of future research needs regarding the potential need for implementation of such incentive mechanism. Some brainstorming and discussion sessions with internal and external IT and ATM experts are planned to support the development of this project objective. These discussions will qualitatively evaluate a set of governance rules that could be implemented through the use of smart contracts and digital tokens.

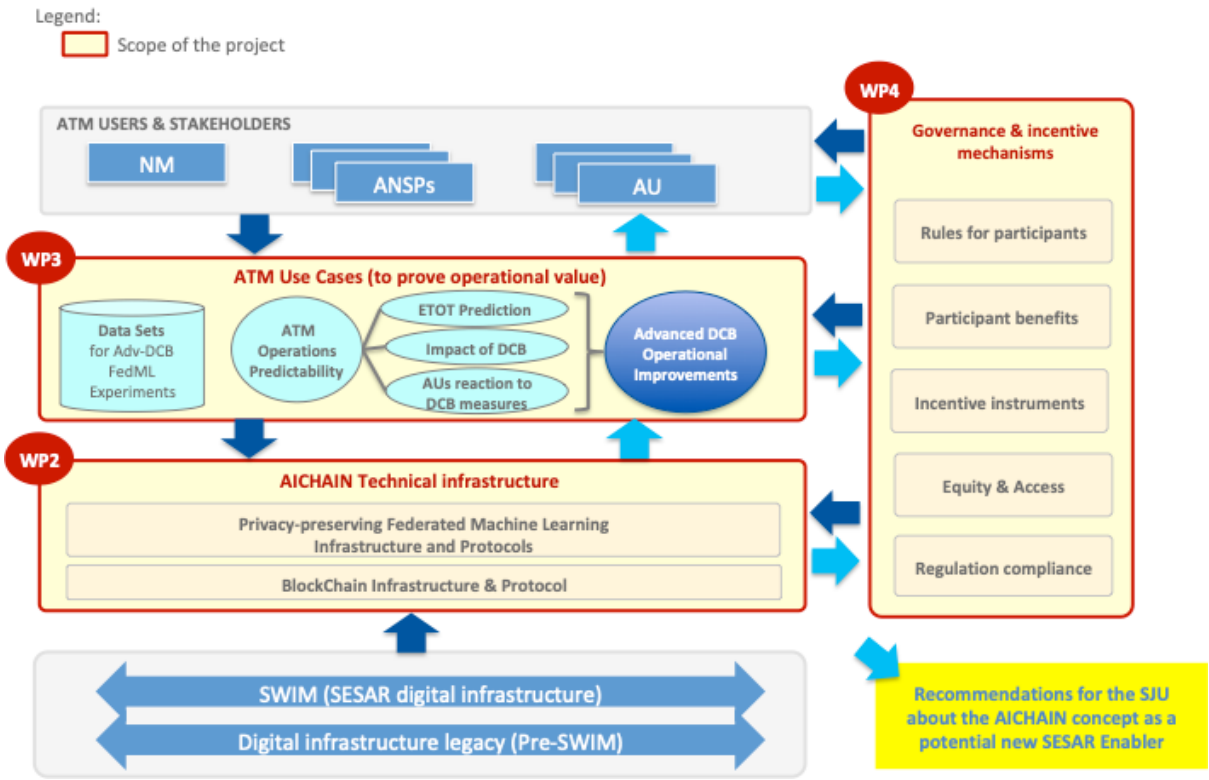


Figure 7-1 Project scope and content of work packages

7.3 Specific objectives per work package/area of research

The <u>specific objectives</u> of the research are structured in three research areas to be covered with different levels of deepness: i) the DIM **technological** solution, ii) the **operational value** of the DIM solution, and iii) the **governance & incentives** aspects. The following are the specific objectives in these three areas:

- **OBJECTIVE #1 (O1):** In the **technological dimension (WP2)**, to define the **AICHAIN Solution architecture** as a potential SESAR technology enabler for the exploitation of private data value, and to implement a **functional small-scale prototype** for user validation and operational value experimentation.

- **OBJECTIVE #2 (O2):** In the **operational dimension (WP3)**, to demonstrate and quantify the operational value of the AICHAIN concept with two **ATM use cases** in the area of A-DCB services.

.

- **OBJECTIVE #3 (O3):** In the **governance dimension (WP4)**, to develop an **incentive mechanism** that addresses the motivational aspects of the data owners in order to facilitate the adoption and the effective utilisation of the AICHAIN concept.

7.4 Research questions and means of verification

The research questions (RQs) addressed in the project, together with the identified means of verification (testability), are presented in Figure 7-2 and further detailed in Table 7-1. The RQs are related to each addressed research area (or work packages), thus following a logical argument line of the research.

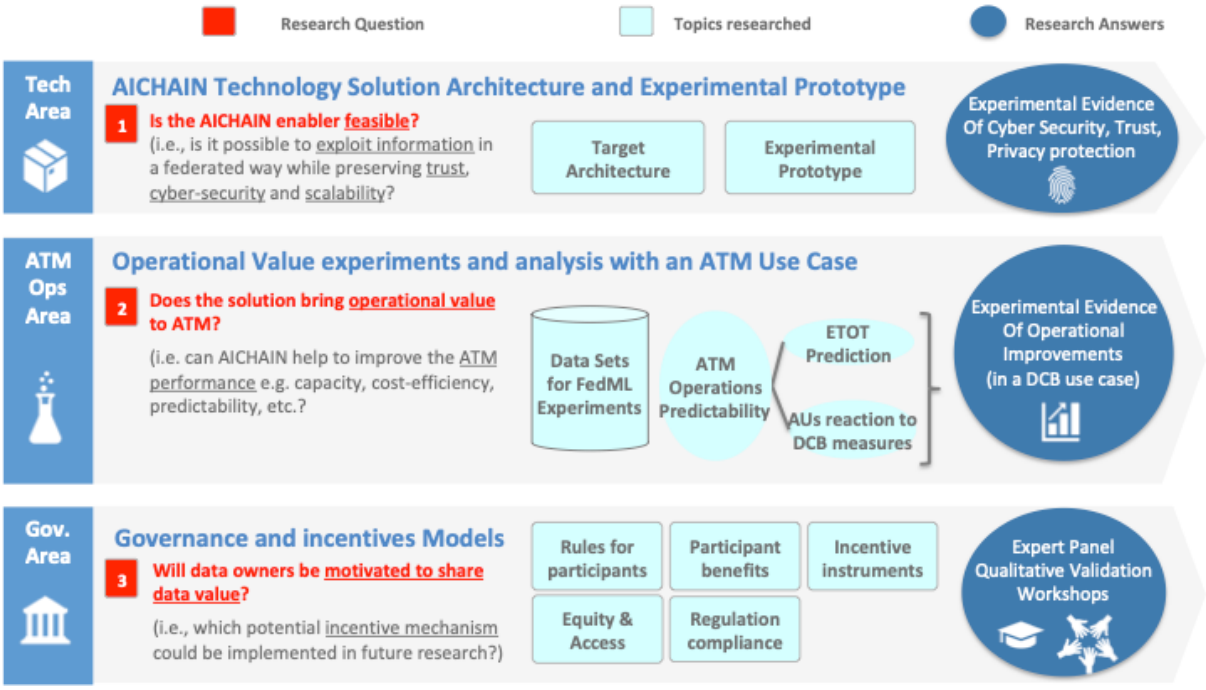


Figure 7-2 Research questions and their testability per area of research/work package

Table 7-1 Research questions and their testability through means of verification and validation

RQ #	Research Questions	Means of verification & validation (= Expected project outcomes)
RQ1	**Is the AICHAIN enabler feasible?** *I.e., is it possible to exploit information in a federated and decentralized way while preserving trust, cyber-security and scalability?* Technological perspective. (around 40% of the project effort)	Architectural analysis
		AICHAIN functional prototype
		Verification of cyber-security data requirements (privacy, integrity, provenance, trust, etc.)
		Verification of big data exploitation requirements (scalability and runtime)
		Use demonstration in an ATM Use Case Experimentation
RQ2	**Does the AICHAIN solution bring operational value to the ATM?** *I.e. Can AICHAIN help to improve the ATM performance e.g. capacity, cost-efficiency, predictability, etc.?* Operational value perspective; (around 40% of the project effort)	Develop ATM Use Cases for Experimentation
		BIM analysis
		Quantify the operational value
RQ3	**How the data owners can be encouraged to collaborate in the AICHAIN concept?** *I.e., Which potential incentive mechanism could be implemented in future research?)* Governance and incentives perspective; (around 20% of the project effort)	Assessment and confirmation of the need for an incentive mechanism
		Development of a governance & incentive scheme
		Qualitative validation of the model in workshops with operational experts.

7.5 Types of evidence that will be generated within the project

The validation plan consists of three main **types of validations**:

- The **Experimental Plan** based on the small-scale experimental prototype (WP2) and the ATM research use cases (WP3) (note: the experiments E1 to E6 are explained in detail later in this document, in the section *8. Experimental Plan and methodology*):

- E1: private data value (standalone experiment / non-federated)
- E2 & E3: (federated experiments with the AICHAIN prototype)
- E4 & E5: cyber-security and privacy-preservation at inference time
- E6*: extension of E3 with additional AU´s (ambitioned but out-of-scope)

- The **external/independent consulting tasks** (outsourcing/procurement Budget)
  - Cyber-security penetration test by independent experts at SWISS IT lab.
  - Regulatory compliance aspects related to data-privacy regulatory.
- The **Expert Advisory Board workshops** (EAB1 and EAB2)
  - Governance & incentives: identification of needs and discussion on concrete proposals
  - General assessment of the ACHAIN concept and its value for ATM

The experiments will generate both **quantitative & qualitative evidence**.

The consulting tasks and the workshops will provide only **qualitative evidence**.

7.6 General validation methodology

The purpose of the research methodology is to establish the means and sequential high-level steps to generate the due quantitative and qualitative evidence for answering the proposed research questions in the most comprehensive and scientifically correct way. In a nutshell, the research methodological steps will follow **three parallel but interdependent workflows** supporting the three aspects of the DIM that have been scoped in the project, i.e.: **technical, operational, and governance & incentives**. The methodology for each of the research workflows is explained in more detail below.

7.6.1 Research methodology for the technological dimension of the DIM

During the definition phase, the **state of the art** of the AICHAIN solution and proposed architecture will be revisited to obtain the most updated insights. In parallel, internal partnership discussions between the technological experts and the operational ones will serve to **identify and elicit the operational requirements** of the AICHAIN enabler. The requirements will be complemented with the **feedback received from a workshop** with external ATM and IT experts. A DIM architecture will be preliminarily selected for the study.

After a **design phase**, a functional **prototype** of the AICHAIN enabler will be developed. Once the prototype will be ready, it will be **used as a demonstrator**. At this stage, a first **verification iteration** cycle will be performed **to test the use case data models and data**

interfaces with the AI models developed to perform the experiments and **using synthetic-data** that will be generated for that purpose.

Further **tests and analyses** will be also applied to the technical infrastructure to **verify whether the DIM requirements are fulfilled** (trust, cyber-security, integrity, etc.). The results of the technical audit will be presented and discussed with stakeholders in a **second workshop at the end of the project**.

7.6.2 Research methodology to assess the potential operational improvements in ATM

To **demonstrate the value and feasibility** of the AICHAIN concept in ATM, the research will explore two **ATM operational use cases** in which the data exploitation enabled by the AICHAIN solution can potentially bring significant operational value. For that purpose the **identification and selection of the most convenient use case** will be based and justified after conducting a state of the art analysis. Together with that, a **preliminary benefit-impact mechanism (BIM)** will be elaborated and altogether discussed with stakeholders in the first workshop.

After the definition phase, the **design of the ML models** and the **design of the experiments** have been conducted.

Before the experiments can be conducted with the prototype, the **data preparation is needed**. Ideally, the **experiments should be setup such that Airspace Users participating in the AICHAIN solution could train and send back a ML federated sub-model based on relevant and possibly sensitive information**, e.g., passenger info, actual fuel and mass, cost index, preferences on winds and weather, preferences on taxation/routes, etc. Note that these pieces of information are typically not visible to NM, thus they could be potentially exploited through the AICHAIN architecture while the privacy and trust are protected; other private data that could be exploited to derive relevant features of the ML models will be identified during the research.

Since involving a large number of AUs in this project is not possible due to resource limitation, the experiment plan will be based on data provided by SWISS conveniently anonymized or under non-disclosure agreements for the purpose of this research (i.e. to be consumed by local machine learning algorithms). This should allow to assess the potential gains in the prediction performance of the ML models with high level of confidence, but the results being valid only for the case of SWISS (i.e., not generalizable).

A **preliminary quantification of the potential benefits** that AICHAIN could bring to the ATM will be conducted through simulations and expert judgment in workshops opened to any airspace users and operational experts. Whenever possible, the results of the experiment will be **compared against the ones obtained through traditional ML techniques that are today being used by NM in current operations and researches,** in which the private data is not available for the training of the AI models (e.g., EUROCONTROL is currently working in the prediction of ETOTs using ML techniques and NM data only). It must be noted that these types of evidence will be preliminary but appropriate to the level of maturity of the AICHAIN solution (according to E-OCVM reference SESAR research pipeline methodology), and thus they should be helpful to

make a go/no-go decision after the delivery of the outcomes of this project with regards to the incorporation of AICHAIN as a potential new SESAR enabler deserving further R&D efforts.

7.6.3 Research methodology for the dimension of governance and incentives mechanism

To effectively generate operational value, the AICHAIN solution depends on the honest collaboration of the data owners (in our use cases, Airspace Users), which is subject to potential "motivational barriers" that must be early assessed. This aspect of the project possibly is the one with lower level of maturity and lack of understanding due to the novelty of the proposed solution. Thus, **the research evidence in this track will be generated mostly through qualitative analysis techniques.**

A **literature review** of the potential incentives mechanisms in ATM or in other domains will be helpful to identify the actual needs, combined with **brainstorming sessions** to preliminarily identify and propose a set of rules, rewards and penalties to encourage stakeholders to participate in the federated training of the models and to do it honestly.

The organisation of **internal discussions** and **external debates in the first workshop** will contribute to assess whether there is a need in the ATM domain to establish an incentive mechanism. **Mathematical analysis** and **economic theory** together with associated **regulatory and political aspects** (e.g. equity and fairness) will be taken into consideration during the design and development of the incentives mechanism. It is relevant to mention here the extensive experience of NOMMON and EUROCONTROL regarding the development of market-based mechanisms and incentive alignment mechanisms in the context of SESAR PJ07 (e.g., the UDPP or the Absolute Priority concepts).

The proposed incentive mechanism will also consider how it could be integrated in the AICHAIN solution through the foreseen utilisation of **smart contracts** and **token schemes** available in the Blockchain toolset. **No implementation or empirical demonstrations of the mechanism will be conducted in this project.**

Once the AICHAIN prototype will be available, and the use case studied, a **second round of concept review**, **brainstorming sessions** in a **second workshop** will help to refine and further develop the governance and incentive mechanism aspects.

7.6.4 Main Assumptions

At this stage of the research, the major assumptions are:

1) The ATM stakeholders are assumed **to cooperate but not to trust each other**.

2) During the ATM use case experiment, **Airspace Users will be assumed to act as "unconditionally honest participants"**, i.e. they always participate and they do it honestly, even in the absence of an incentive mechanism.

7.7 Summary of the validation plan (view per work package)

Table 7-2 shows a summary of the main means of verification and validation explained so far, mapped to each of the work packages / areas of research / research questions of the project. Note that experiments E1 to E6 will be explained in detail in next section.

Table 7-2 Means of testability seen per work package / area of research

WP (RQ) #	Item to validate	Means of verification/validation
WP2 (RQ1)	Feasibility of the federated learning approach	• Architectural analysis • Exp. E2, E3 (with the prototype)
	Cyber-security and privacy-preservation	• Exp. E4, E5 • SWISS assessment (by their IT experts) • Penetration tests and expert assessment (outsourced) • Literature and technology market assessment
	Scalability	• Exp.E3 & E4 • Exp.E6 (–out-of-scope-) • (Scaleout background tests published in scientific journals)
WP3 (RQ2)	ATM value/performance	• Exp. E1, E2/E3, E6* (*note: E6 is identified as relevant but it is not in the scope of this project)
WP4 (RQ3)	Governance and incentives	• Rules and incentives mechanism: Literature review + expert assessment (EAB workshops) • Regulatory compliance: Expert assessment (outsourced)

8 Experimental plan and methodology

To **assess the potential benefits for the ATM system** that the AICHAIN Technological Solution could bring, both **quantitative and qualitative evidence** will be gathered **through the execution of experiments with a prototype and through expert judgment** validation in workshops. Whenever possible, the results of the experiments will be compared against the ones obtained through traditional ML techniques that are today used by NM in operations, in which the private data is not available for the training of the AI models (e.g., EUROCONTROL is currently working in the prediction of ETOTs using ML techniques and NM data only). It must be noted that these types of evidence are considered **appropriate to the level of maturity** of the AICHAIN solution (according to E-OCVM reference SESAR research pipeline methodology), and thus they should be helpful to make a go/no-go decision after the delivery of the outcomes of this project with regards to the incorporation of AICHAIN as a potential new SESAR enabler deserving further R&D efforts.

Since involving a large number of airspace users in this project is not possible due to resource limitations, the experiments plan will be based on **data provided by SWISS only** (in addition to the NM data). The SWISS dataset will be also partitioned (horizontally) to feed different federated nodes, emulating the case that several different airspace users are participating in the federated alliance.

8.1 Experimental setup

8.1.1 ML model architectures and interface with the AICHAIN prototype

The AICHAIN prototype has been designed to be interoperable with many different and most common machine learning platforms available in the market and to facilitate in extremis the federation of the training and inference processes of machine learning models.

This means that there is no need to develop any special or different code for the federated experiments. The same Python code can be reused for both the non-federated and federated experiments, since the FEDn module is in charge of coordinating the training and prediction processes with the aggregation node acting in synchronisation with the rest of federated alliance.

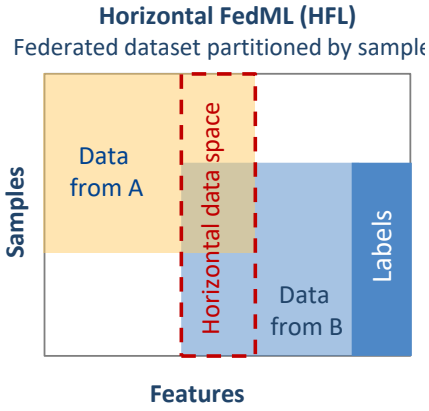
The only constraint at this moment is that the ML model used must be a **neural network**, but in the future other ML model architectures will be also possible.

8.1.2 Federated setup and actors (physical and logical architectures)

In Federated Machine Learning the federation can be classified according to how the datasets are partitioned between the nodes with respect the shared ML model to be trained. See Figure 8-1. These aspects and others related to the federated machine learning technology are explained in deep detail in Deliverable D2.1.

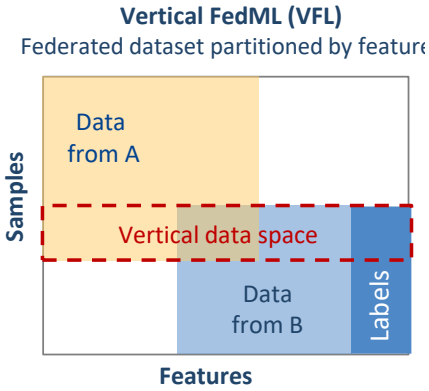
- **Horizontal Federated Learning (HFL)**
  Extended sample space from partners and a common intersected feature set.

  In HFL the data is partitioned horizontally, or by sample, i.e. each participant has different samples but the feature space is the same between different participants. For example, similar scientific instruments that are owned by different companies that train a shared model with their specific instrumental data.



- **Vertical Federated Learning (VFL)**
  Extended feature space from partners and a common sample space.

  In VFL the data is partitioned vertically, or by features, i.e. different participants might hold different types of information about the same example. For instance, a bank or an insurance company that hold different types of data about the same customer.



- **Federated Transfer Learning (FTL),**

  In FTL the datasets may differ in sample and/or feature space. It is used when the FedML participants have very heterogeneous datasets, and some of the nodes don't have enough samples or features. The essence of transfer learning is to find the invariant between a rich-data source and a data-scarce target, and exploit that invariant to transfer knowledge from the source domain to the target domain.

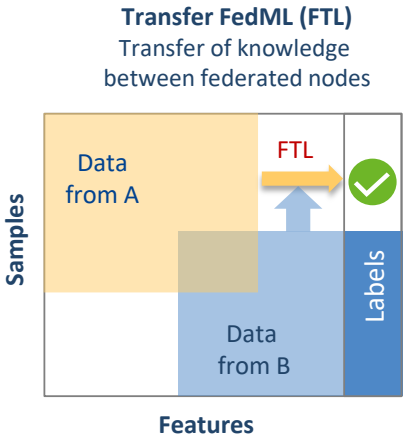


Figure 8-1 Federated Learning types by Data partition (figs adapted from [24])

This project only has available data from two actors: EUROCONTROL and SWISS. Since each AU has all the information for their own flights, horizontal federation may be performed, since NM data is shared with the AUs. Hence, flight-specific data will be merged with shared NM data in the AUs' data silos in an agreed manner so it can be used in the training of the federated model. For practical reasons, the two datasets available will be concatenated and partitioned horizontally every 2 or 3

months to emulate more than one AU in the federated alliance. Therefore, all the datasets in the federated setup will have different samples but the same feature structure.

8.1.3 Workflow

Figure 8-2 schematises the information exchange workflow between the nodes having the private data and the orchestrator. Note that the information exchange is not raw data but the privacy-preserving information that contains the model parameters to be aggregated in the orchestrator.

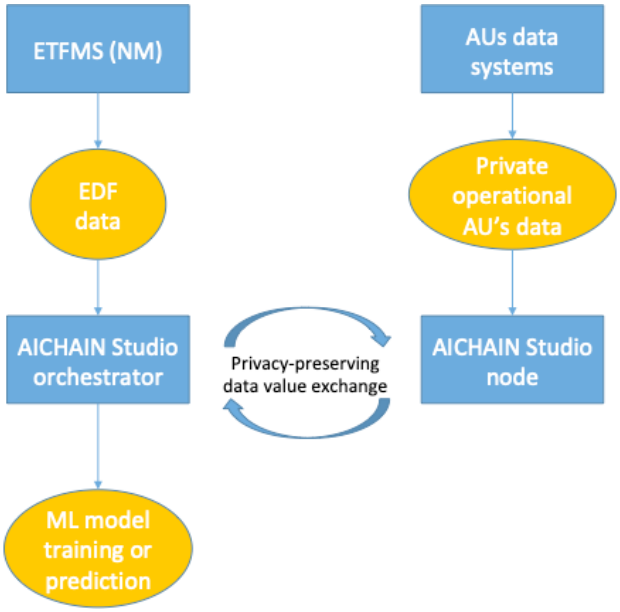


Figure 8-2 Workflow of information exchanges between the nodes and the orchestrator

8.1.4 Dataset description

The following table shows the description of the datasets to be used in the experiments.

Table 8-1 Dataset description for the UC1 and UC2

UC	NM dataset	SWISS dataset
1	• Flights: All flights crossing MUAC sectors • Period: from 27/06/2019 to 28/02/2020 • Source: ETFMS • Samples: 68.2M (from which 1.07M correspond to SWISS flights) • Features: see Section 5.2	• Flights: All flights crossing MUAC sectors • Period: from 27/06/2019 to 28/02/2020 • Source: SWISS • Samples: 4,8 M • Features: see chapters 5.3

2	• Flights: All flights connecting Switzerland and the great London area • Period: from 27/06/2019 to 28/02/2020 • Source: DDR-2, ECMW • Samples: 20K (from which 10K correspond to SWISS flights) • Features: see Section 6.2	• Flights: All SWISS flights connecting Switzerland and the great London area • Period: from 27/06/2019 to 28/02/2020 • Source: SWISS • Samples: 10K • Features: see Section 6.3

8.2 Design of experiments

The purpose of the experiments in WP3 is to assess the value of the private data that normally is not accessible to NM but that could be available for training ML models through the AICHAIN solution. The main goal of these experiments is to test the value of private data to improve ATM operations models' performance

Table 8-2 identifies the set of experiments to be conducted together with a high-level description of the goals and rationale, and indicating the WP context that is closer to the type of evidence. These experiments are further described with more detail in the following subsections.

Table 8-2 Description of the experiments to assess the AICHAIN Technological Solution

E#	Description	Objectives and rationale	WP context
E1	**Benchmark in standalone mode of the use case model improvement with private data**	• To assess the **value of private data (non-federated).** • To **set the benchmark reference** for E2	**WP3 (RQ2)**
E2	**Functional test of the federated learning platform and comparing to the baseline E1**	• To test the **technical feasibility** of AICHAIN prototype • To **benchmark against E1**: has the federated model lose **accuracy** ? • To **set the benchmark reference for E3** (communication-computation aspects)	**WP2 +WP3** **(RQ1&RQ2)**
E3	**Full platform test with FEDn Client at Swiss secured host environment**	• To test the **technical feasibility (and somehow operational)** of the AICHAIN through heterogeneous and remote IT systems (server at SITA and client at	**WP2 +WP3** **(RQ1&RQ2)**

		SWISS) • **To benchmark against E2**: has the federated model lose **accuracy** ? • **To benchmark against E2**: assess communication-computation performance, cyber-security and privacy preservation characteristics	
E4	**Test of model provenance tracking with Blockchain**	• To test the **technical feasibility of the Blockchain components** • To assess the **performance overhead vs cyber-security** (*trust)* gains • To assess the **privacy gains** by Blockchain enablement (BLOCKn), by means of mechanisms (**Provenance, Traceability, Auditability**)	**WP2 (RQ1)**
E5	**Test of model serving and inference**	• To assess the **cyber-security aspects** at model serving/inference time • To assess the **governance aspects** of federated consumption .	**WP2 (RQ1)**
E6	**Test with additional airlines (ambitioned but out-of-scope)**	• To test **technical feasibility** with more heterogeneous systems • To **engage additional airlines (exploitation)** • To **gather feedback** about pros and cons of the technology • To **benchmark against E3 results**	**RQ1&RQ2** **(identified as relevant but out of scope)**

8.2.1 Experiment E1

The experiment E1 (*"benchmark in standalone mode of the use case model improvement with private data"*) aims at finding the value of the private dataset in the case that this data would be directly available to NM without the need for federating the ML model training.

Figure 8-3 shows the general methodology for benchmarking the different experiments that will be conducted as part of the research. The experiment consists in comparing a baseline (or reference) scenario (model V0) against the benchmark (or solution) scenario (model V1).

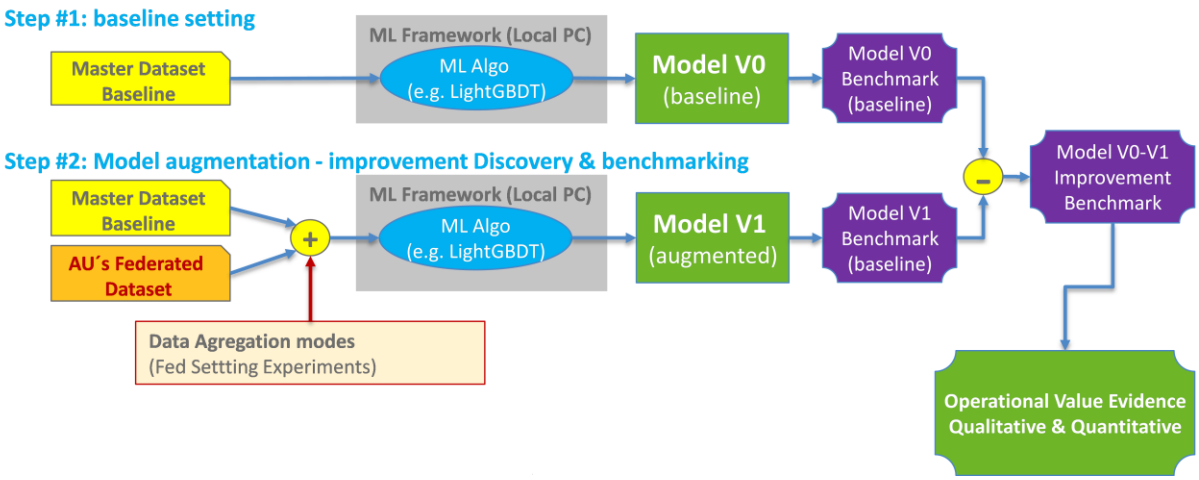


Figure 8-3 Methodology and setup for benchmarking scenarios/models in experiment E1

Table 8-3 describes the baseline and solution scenarios that will be generated for each use-case (UC1 and UC2). Since only private data from SWISS is available, two baseline scenarios have been considered. The rationale is explained in the table. Note also that two performance indications will be gathered from scenario B0, one that will measure the APELAT (Average prediction (absolute) error as a function of the LAT (look-ahead time)) for all the flights and the other by specific for the SWISS flights (measure from the model trained with all the flights).

Table 8-3 Description of the baseline and solution scenarios

Id	Training dataset	Test dataset	Performance indicator(s)	Rationale
B0	Flights: All flights Samples: NM Features: NM # samples: 68.2M**	Randomly chosen from the UC1 training dataset	1) APELAT*, for all the flights 2) APELAT*, for SWISS flights	Comparison of the SWISS APELAT with later experiments can help to form a qualitative intuition of the potential improvement if the private datasets of all the AUs were available.
B1	Flights: SWISS flights Samples: NM Features: NM # samples: 1.07M**	Same as in B0	1) APELAT*, for SWISS flights	The results of this experiment will be the actual baseline to assess the solution scenario, i.e., to assess the added value of SWISS data in the prediction model.
S1	Flights: SWISS flights Samples: NM Features: NM + SWISS # samples: 1.07M**	Same as in B0	1) APELAT*, for SWISS flights	The solution scenario assumes that NM samples can be enriched with SWISS features by horizontal federation.

* Average prediction (absolute) error in minutes, as a function of the prediction LAT (look-ahead time)

** The final number of samples for UC1, UC2 dataset are still under construction

8.2.2 Experiment E2

The experiment E2 (*"Functional test of the federated learning platform and comparing to the standalone baseline")* aims at finding the value of the private dataset in the case where the private dataset is available through federation using the AICHAIN prototype. In this experiment both the master and partner nodes will be placed in the same cluster, thus avoiding the complexities of performing the federated ML training with remote systems interconnected through the Internet.

Figure 8-4 shows the experimental methodology and setup for this experiment. The experiment consists in comparing the scenario S1 (model V1) from experiment E1 with the new scenario S1* (model V1*) generated with the E2 setup. The major objective is to observe if there is a degradation of the model performance due to the application of federation techniques. For practical reasons, the two datasets available will be concatenated and partitioned horizontally every 2 or 3 months to emulate more than one AU in the federated alliance. Therefore, all the datasets in the federated setup will be composed of different samples but the same feature structure.

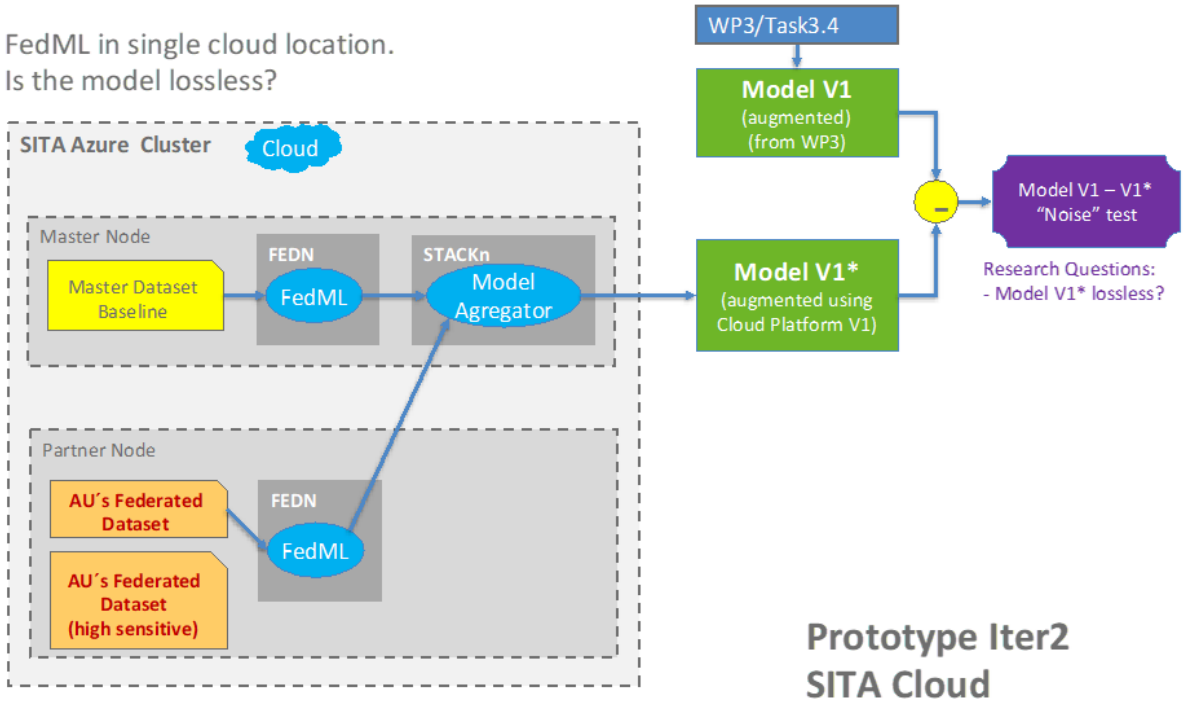


Figure 8-4 Methodology and setup for experiment E2

8.2.3 Experiment E3

The experiment E3 (*"Full platform test with FEDn Client at Swiss secured host environment")* is similar to E2 with the major change being that the master node and the federated node will be remotely separated and interconnected through the Internet. The master will be at SITA facilities, representing the NM, and the federated node will at SWISS facilities, representing to itself (realistic interconnection and interoperable setup).

Figure 8-5 shows the experimental methodology and setup for this experiment. The experiment consists in assessing the feasibility and runtime performance aspects, while double-checking that the model performance is the same as in E2. For practical reasons, the two datasets available will be concatenated and partitioned horizontally every 2 or 3 months to emulate more than one AU in the federated alliance. Therefore, all the datasets in the federated setup will have different samples but the same feature structure.

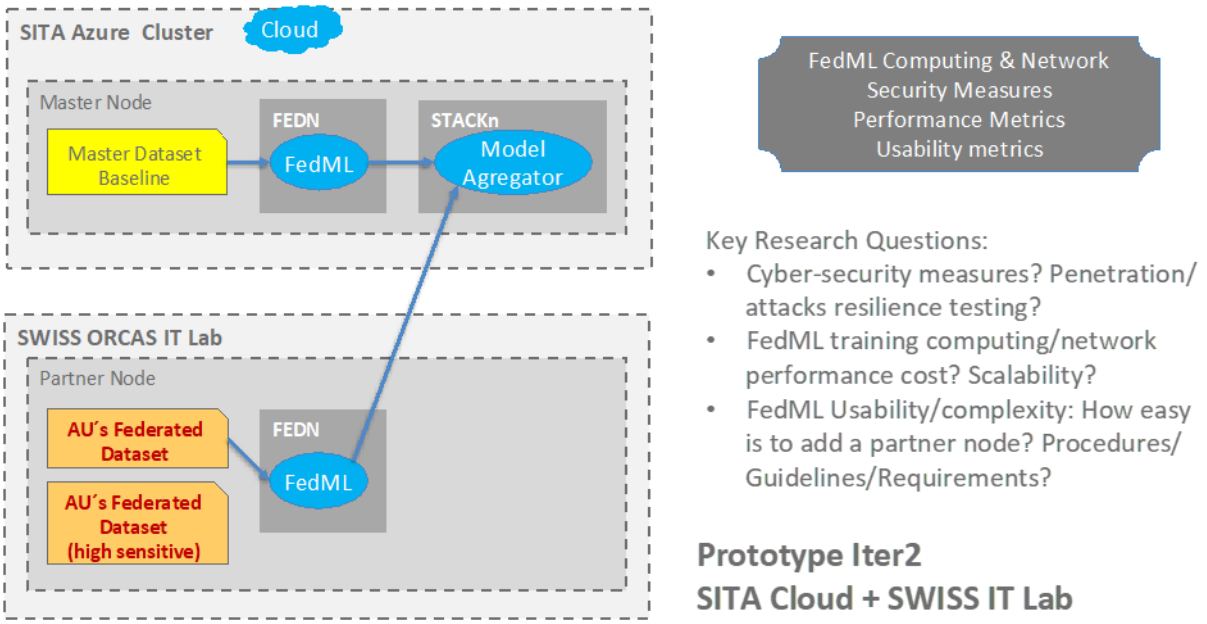


Figure 8-5 Methodology and setup for experiment E3

8.2.4 Experiment E4: Test of model provenance tracking with Blockchain

Experiment E4 will be further described in future deliverable D2.2 (WP2), as soon as all of the blockchain components of the architecture will be ready (now under development). For the sake of completeness, the following high-level description of the experiment is provided.

Purpose

- To use the Blockchain components to provide provenance tracking the model training, as a technique to support privacy-preserving features of federated training.
- To assess the performance overhead vs cyber-security (provenance for *trust)* gains
- To assess the privacy gains by Blockchain enablement (BLOCKn), by means of mechanisms
    - Provenance
    - Traceability
    - Auditability

8.2.5 Experiment E5: Test of model serving and inference

Experiment E5 will be further described in future deliverable D2.2 (WP2), as soon as all of the blockchain components of the architecture will be ready (now under development). For the sake of completeness, the following high-level description of the experiment is provided.

Purpose

- The intended use of federated ML and Blockchain components to allow for tailored governance aspects of federated consumption.

Elements that will be assessed:

- Cyber-security during model Inference
- Governance Model of inferences request ratio, and possible use of smart-contract inference based protocol

8.2.6 Experiment E6 (out of the project scope)

The experiment E6 (*"Full platform test with FEDn Client at Swiss secured host environment")* is similar to E3 but including several AUs as data feeders in the federated alliance. The purpose is to assess the feasibility and runtime performance, while assessing how much the model accuracy could be improved by adding more data feeders.

Figure 8-6 shows the experimental methodology and setup for this experiment. Due to the budget and temporal limitations of this project, this experiment is out of the project scope and will be proposed as continuation for further research.

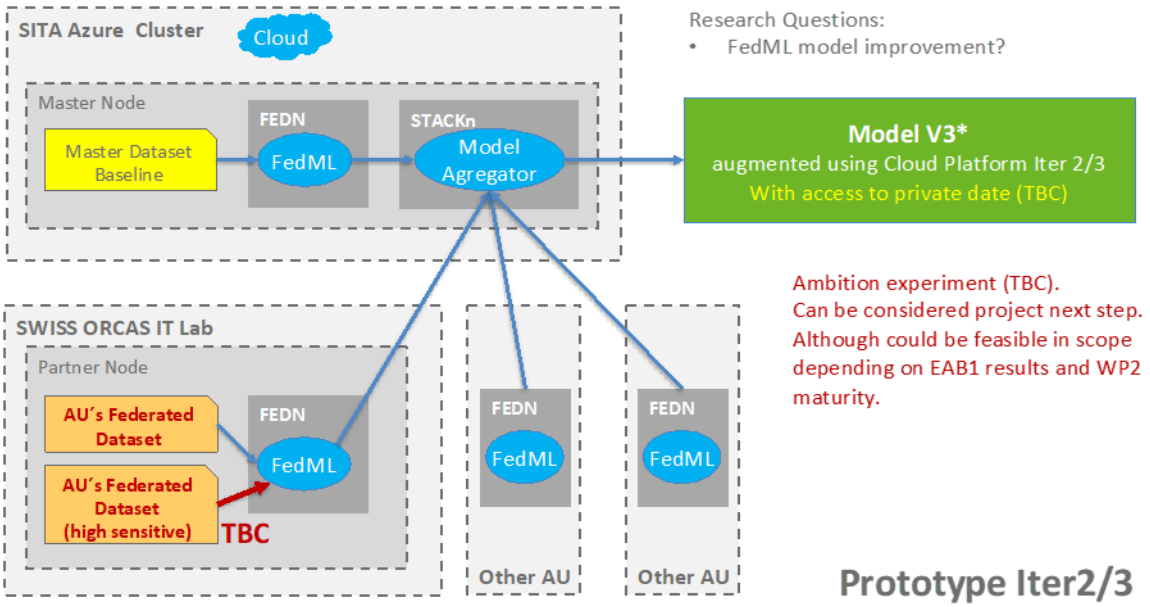


Figure 8-6 Methodology and setup for experiment E6

9 Preliminary experimentation and proof of value for UC1

9.1 Description of the experiments performed and scope

This section presents the preliminary results of some experiments already conducted with an early (non-stable) version of the UC1 model. The experiments will be further assessed in the second year of the project, and complemented with the full battery of experiments described in the previous chapter. The present results are useful to have an anticipated evidence of the potential value of the AICHAIN Technological Solution for ATM.

The experiment conducted is the experiment E1 and the scenarios generated are the ones identified in previous section of the document.

9.2 Description of dataset partitioning and training results

The available dataset was split in three different sets for the experiment:

- **Training set:** from 2019-06-27 to 2019-10-31, which corresponds to 807687 samples obtained from 49155 flights.
- **Validation set:** from 2019-11-01 to 2019-12-31, which corresponds to 231531 samples obtained from 18727 flights.
- **Test set:** from 2020-01-01 to 2020-02-28, which corresponds to 228666 samples obtained from 17977 flights.

Figure 9-1 shows the learning curves of scenarios B1 and S1 respectively.

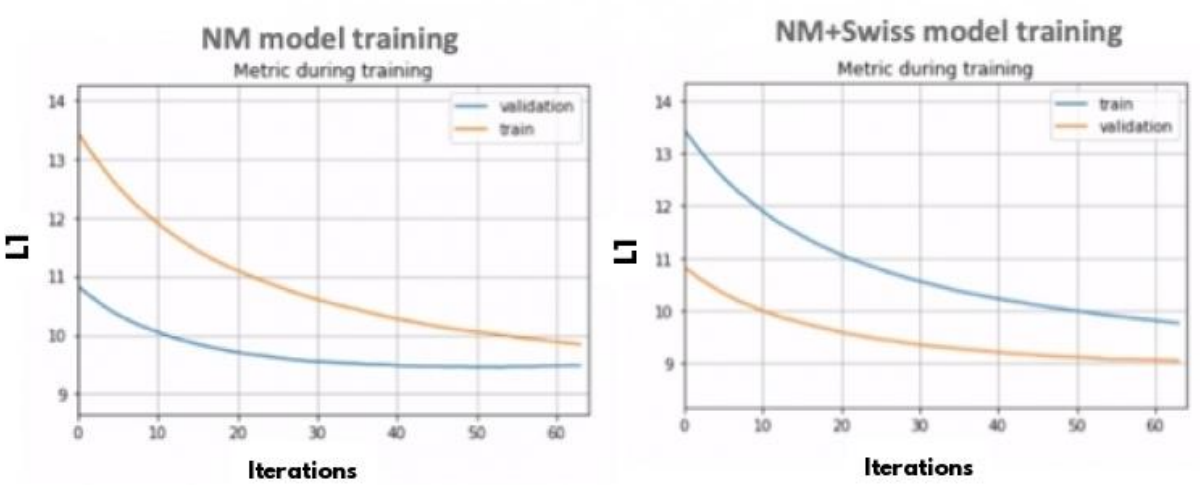


Figure 9-1 Learning curves (mean absolute error, also known as L1, as a function of the training steps)

9.3 Results

9.3.1 Model performance results

Table 9-1 shows the results in absolute terms for the 3 scenarios B0, B1, S1 (the results of B0 are shown for the case of ALL the flights and for the flights of SWISS only). Table 9-2 shows the results in relative terms with respect to the scenario B1. The following observations can be done:

- The flights of SWISS are more predictable than the average of all the flights.
- The scenario S1 is better in terms of predictability than B1, meaning that the private data of SWISS has significant value to increase the predictability.
- The scenario B0_SWISS presents better predictability than B1, which highlights the importance of having large amounts of data to increase the predictability of the SWISS flights, even if the samples are about other airspace users´ flights. In this case it can be observed that more data for training the ML model (scenario B0_SWISS) is more important that more data quality / extra features (scenario S1).

Table 9-1 Absolute mean error as a function of the look-ahead-time in absolute terms

Scenario / LAT	(0, 30]	(30, 60]	(60, 90]	(90, 120]	(120, inf]
B0_ALL	6.45	9.61	10.75	10.83	14.09
B0_SWISS	4.92	7.62	8.76	9.09	10.90
B1	5.12	7.87	9.16	9.47	11.19
S1	5.06	7.70	8.94	9.19	11.04

Table 9-2 Mean error relative to scenario B1 as a function of the look-ahead-time in relative terms

Scenario / LAT	(0, 30]	(30, 60]	(60, 90]	(90, 120]	(120, inf]
B0_ALL	25,98%	22,11%	17,36%	14,36%	25,92%
B0_SWISS	-3,91%	-3,18%	-4,37%	-4,01%	-2,59%
B1	0,00%	0,00%	0,00%	0,00%	0,00%
S1	-1,17%	-2,16%	-2,40%	-2,96%	-1,34%

Table 9-3 Absolute mean error of the predictions on the test set

	ETFMS	B1	S1
Mean	10.5	9.3	9.0
Std deviation	15.7	13.0	13.0
Q1	2.9	2.7	2.6
Q2	6.0	6.1	5.8
Q3	12.0	11.4	10.8

Table 9-3 shows the absolute error on average for the ETFMS predictions, B1 and S1. The following results can be inferred:

- Improvement with NM data only (scenario B1) with respect the ETFMS predictions: 11.4%
- Improvement with NM+SWISS data (scenario S1) with respect the ETFMS predictions: 14.29%
- Relative improvement ((ETFMS-S1)/(ETFMS-B1)-1): +25%

9.3.2 Feature importance analysis

It is interesting to analyse the importance of the features on the predictability performance of the model. Figure 9-2 shows the importance feature analysis for B1 (left) and S1 (right). In the right figure the most important features provided by SWISS as part of its private data have been highlighted in red. Among the features that contribute more to increase the predictability of the take-off time of the SWISS flights we can find the number of passengers booked and on board, the runway assigned, the own predictions of SWISS of the ETOT, the assigned gate, the number of passenger with connection and the time of connection of the previous flight.

In future analyses an enhanced and more detailed analysis will be conducted to assess the contributions of the different private features.

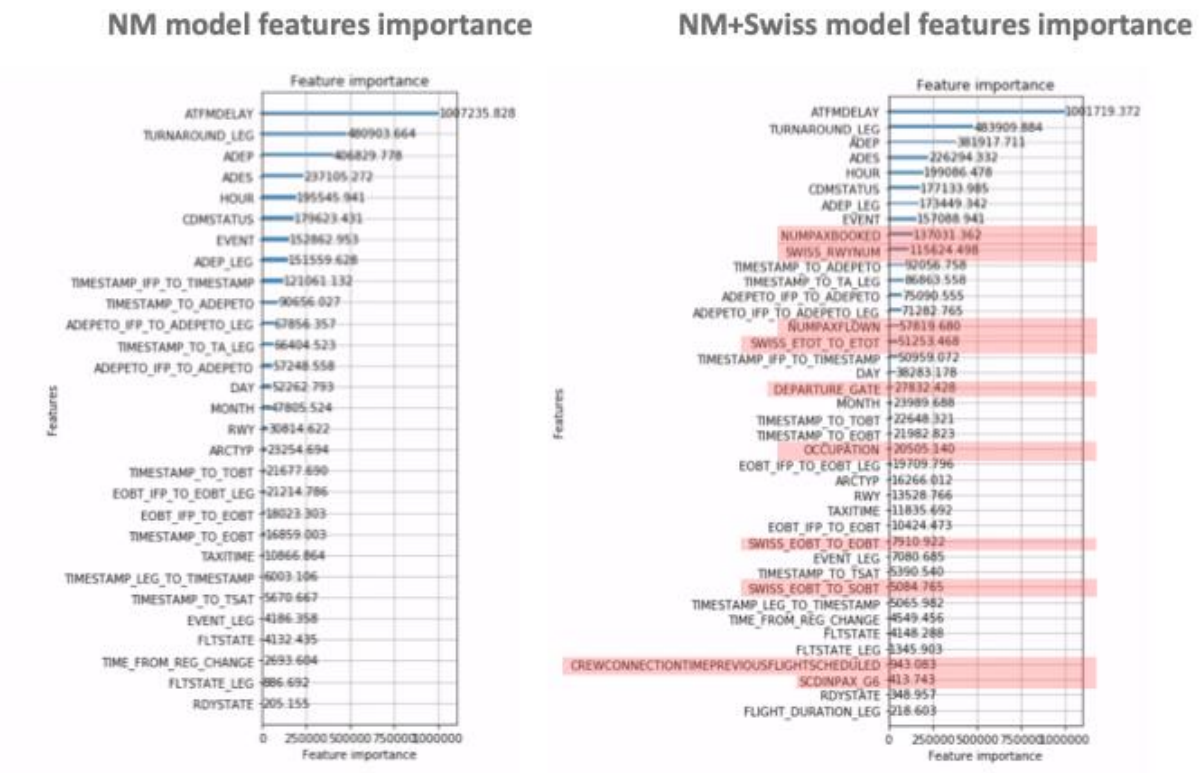


Figure 9-2 Feature importance analysis

9.4 Conclusions of the early experiments with UC1

After the analysis of the early evidence generated with UC1 the following conclusions can be extracted:

1) The exploitation of **SWISS private data could improve the ML model predictions.**

2) The experiments were conducted with **relatively low volume of data** to train the models with SWISS flights only. Therefore, **more AUs** are needed to achieve more representative results.

3) As in the non-federated ML models, just adding more data is not enough. A **good depuration of the input and a fine-tuning of the model is required** to achieve good performances.

4) This use case demonstrated the value of data to improve the ML models, but we cannot say it could have a significant positive impact on ATM operations (**proof of value for ATM operations could not be shown**).

10 Interim conclusions about the operational value of AICHAIN

In this interim deliverable of WP3 the following evidence about the potential value of the AICHAIN Technological Solution for ATM has been generated:

- Qualitative evidence: chapters 2 to 5
  - AICHAIN is well aligned with the current efforts to introduce higher levels of automation based on advanced AI/ML models.
  - The technological solution could solve –if it is proven feasible– the handicaps identified linked to 'data sharing', which cannot work well when the private data is subject to certain privacy requirements.
  - AICHAIN is in line with all the strategic documents of the future ATM, e.g. the ATM Master Plan, the Aviation Strategy 2050, the Airspace Architecture Study, the Digital European Sky, the Fly AI Report, the report of Automation in ATM, and others. This indicates that AICHAIN could play an important role in the digitalisation process of the ATM.
  - As a technological enabler, AICHAIN is not restricted to any specific use case, and indeed could potentially provide large benefits to the different ML models that will support most of the use cases identified as relevant for the ATM by the FLY AI report.
- Quantitative evidence: preliminary results in chapter 9
  - In this deliverable some early and preliminary results have been included, showing the value of the private data in a particular simplified use-case scenario. A significant improvement has been found of about 25% in the accuracy of the ML model trained with NM and SWISS private data with respect the ML model trained with NM data only. These results will be further validated and tested in a real federated setup and will be analysed with more detail and with specific tools for federated environments.
  - Results suggested that the exploitation of multiple sources of private data (i.e. not only SWISS data as in these experiments) and with significantly more quantity of data at each node (e.g. big data) should significantly improve the ML model performance. Such type of more realistic experiments is out of the scope of this project but it is clearly identified a way forward to further assess the operational value of this promising technology.

Further evidence and conclusions will be generated in next deliverable D3.2.

Appendix A SotA article and literature references

[1] SESAR Joint Undertaking, 2020, "The Digital European Sky, Blueprint", doi:10.2829/61772

[2] SESAR Joint Undertaking, 2020, "EUROPEAN ATM MASTER PLAN", doi:10.2829/695700

[3] SESAR Joint Undertaking, 2019, "A proposal for the future architecture of the European airspace", doi:10.2829/309090

[4] SESAR Joint Undertaking, 2019, "A proposal for the future architecture of the European airspace".

[5] SESAR Joint Undertaking, 2020, "SESAR Automation in Air Traffic Management".

[6] European Union , 2011. "Flightpath 2050, Europe's Vision for Aviation", Report of the High Level Group on Aviation Research

[7] EUROPEAN AVIATION ARTIFICIAL INTELLIGENCE HIGH LEVEL GROUP. The FLY AI Report, Demystifying and Accelerating AI in Aviation/ATM, 5th March 2020, https://www.eurocontrol.int/press-release/fly-ai-report

[8] EASA Artificial Intelligence Roadmap, A human-centric approach to AI in aviation, V1.0, Feb 2020, https://www.easa.europa.eu/newsroom-and-events/news/easa-artificial-intelligence-roadmap-10-published

[9] Andrienko, G., Andrienko, N., Fuchs, G., & Garcia, J. M. C. (2017). Clustering trajectories by relevant parts for air traffic analysis. *IEEE transactions on visualization and computer graphics*, *24*(1), 34-44

[10] Ayhan, S., & Samet, H. (2016a). Aircraft trajectory prediction made easy with predictive analytics. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 21-30). ACM

[11] Alpaydin, E. (2009). *Introduction to machine learning*. MIT press.

[12] Ayhan, S., & Samet, H. (2016b). Time series clustering of weather observations in predicting climb phase of aircraft trajectories. In *Proceedings of the 9th ACM SIGSPATIAL International Workshop on Computational Transportation Science* (pp. 25-30). ACM

[13] Basora, L., Morio, J., & Mailhot, C. (2017, November). A Trajectory Clustering Framework to Analyse Air Traffic Flows. In *SIDs 2017, 7th SESAR Innovation Days*.

[14] Bian, J., Tian, D., Tang, Y., & Tao, D. (2018). A survey on trajectory clustering analysis. *arXiv preprint arXiv:1802.06971*.

[15] Boucquey, J., Gonze, F., Hately, A., Huens, E., Irvine, R., Steurs, S., & Jungers, R.M. (2017). Probabilistic Traffic Models for Occupancy Counting. *7th SESAR Innovation Days*.

[16] Calvo, E., Cordero, J.M., Vouros, G., Pelekis, N., Kravaris, T., Georgiou, H., Fuchs, G, Andrienko, N., Andrienko, G., Casado, E., Scarlatti, D. & Scarlatti, D. (2017). DART: A Machine-Learning Approach to Trajectory Prediction and Demand-Capacity Balancing. Seventh SESAR Innovation Days.

[17] Delgado, L. (2015, June). European route choice determinants. In *11th USA/Europe Air Traffic Management Research and Development Seminar*

[18] Ghasemi Hamed, M., Gianazza, D., Serrurier, M., & Durand, N. (2013). Statistical prediction of aircraft trajectory: regression methods vs point-mass model. ATM Seminar.. 10th USA/Europe Air Traffic Management Research and Development Seminar, Chicago, United States.

[19] Liu, Y., Hansen, M., Lovell, D. J., & Ball, M. O. (2018, June). Predicting Aircraft Trajectory Choice-A Nominal Route Approach. In *Proceedings of the 8th International Conference on Research in Air Transportation, Barcelona, Spain*

[20] Marcos, R., García-Cantú, O., & Herranz, R. (2018). A Machine Learning Approach to Air Traffic Route Choice Modelling, arXiv:1802.06588.

[21] Naessens, H., T. Philip, M. Piatek, K. Schippers, and R. Parys (2017) Predicting flight routes with a Deep Neural Network in the operational Air Traffic Flow and Capacity Management system. Predicting flight routes February 2018, EUROCONTROL

[22] Valenzuela Romero, A., Franco Espín, A., & Rivas Rivas, D. (2017). Sector demand analysis under meteorological uncertainty. In *7th European Conference for Aeronautics and Space Sciences (EUCASS 2017 Conference)(2017)*

[23] Wang, Z., Liang, M., & Delahaye, D. (2017, November). Short-term 4d trajectory prediction using machine learning methods. In *SID 2017, 7th SESAR Innovation Days*

[24] Yang, Q., Liu, Y., Chen, T., and Tong, Y. Federated machine learning: Concept and applications. ACM Transactions on Intelligent Systems and Technology (TIST), 10(2):1–19, 2019., 13-Feb-2019, arXiv:1902.04885

[25] Dalmau, R., Ballerini, F., Naessens, H., Belkoura, S., and Wangnick. S., 2019. Improving the Predictability of Take-off Times with Machine Learning, Proceedings of SESAR Innovation Days 2019.

[26] EUROCONTROL, 2010. Airport CDM network impact assessment. Technical Report.

[27] EUROCONTROL 2005, AIRPORT CDM Cost Benefit Analysis (CBA). Technical Report.

[28] Tobaruela, G., 2015. A framework to assess the ability of automation to deliver capacity targets in European airspace. PhD Thesis, Imperial College London.

[29] Pickup, S. and Huet, D., 2017. Airport – Collaborative Decision Making (A-CDM) Local and Network Impact Assessment. Proceedings of 12th ATM R&D Seminar

[30] EUROCONTROL, 2021. ATFCM Operations Manual - Edition 25.0. Technical document.

[31] Delgado, L., Gurtner, G., Bolic, T., Castelli, L., 2021. Estimating economic severity of Air Traffic Flow Management regulations. Transportation Research Part C: Emerging Technologies, Vol. 125.

[32] Ruiz, S., 2013. Strategic de-confliction to enable seamless aircraft conflict management. PhD Thesis, Universitat Autònoma de Barcelona.

[33] Lambelho, M., Mitici, M., Pickup, S., Marsden, A., 2020. Assessing strategic flight schedules at an airport using machine learning-based flight delay and cancellation predictions. Journal of Air Transport Management 82,101737

[34] Truong, D., 2021. Using causal machine learning for predicting the risk of flight delays in air transportation. Journal of Air Transport Management91, 101993. doi:10.1016/j.jairtraman.2020.101993

[35] Dalmau, R., Ballerini, F., Naessens, H., Belkoura, S., Wangnick, S., 2021a.An explainable machine learning approach to improve take-off time pre-dictions, p. 102090. doi:10.1016/j.jairtraman.2021.102090

[36] Dalmau, R., Murgese, G., De Wandeler, Y., Correira, R., Marsden, A., 2021a.Early Detection of Night Curfew Infringements by Delay Propagation with Neural Networks, in: 14th USA/Europe Air Traffic Management Research and Development Seminar, New Orleans, LA

END OF DOCUMENT