Issue 24(4), 2024, pp. 18-40
ISSN: 1567-7141
TU Delft OPEN

EJTIR

Special issue
https://ejtir.tudelft.nl
DOI: 10.59490/ejtir.2024.24.4.7517

Roadmap for a European Open Science Alliance for ATM Research

Tatjana Bolić[1], Andrew Cook[2], Rainer Koelle[3], Enrico Spinielli[4], Quinten Goens[5], Martin Strohmeier[6]

[1] corresponding author t.bolic@westminster.ac.uk, Centre for ATM Research, University of Westminster, UK; ORCID 0000-0002-5530-7080
[2] Centre for ATM Research, University of Westminster, UK; ORCID 0000-0002-9144-2343
[3] European Green Sky Directorate, EUROCONTROL, Belgium; ORCID 0000-0003-2810-1404
[4] European Green Sky Directorate, EUROCONTROL, Belgium; ORCID 0000-0001-8584-9131
[5] European Green Sky Directorate, EUROCONTROL, Belgium; ORCID 0000-0003-2187-1924
[6] OpenSky Network, Switzerland; ORCID 0000-0002-1936-0933

Keywords
Open science
Open data
Open code
Open access
Reproducibility
Research community
ATM

Publishing history
Submitted: 16 April 2024
Revised date(s): 27 August 2024
Accepted: 04 November 2024
Published: 08 November 2024

Cite as
Bolić, T., Cook, A., Koelle, R., Spinielli, E., Goens, Q., Strohmeier, M. (2024). Paper Heading. *European Journal of Transport and Infrastructure Research, 24*(4), 18-40.

Abstract

To meet the future challenges of air transport and air traffic management (ATM), higher levels of transparency and access to the underlying data will be a key enabler.

We here propose an Open Science Alliance for ATM, advocating for open data in such a framework. The benefits of adopting an open science approach are to be found, *inter alia*, through the independent verification and validation of reported impacts/results and achieved performance levels, i.e., through reproducibility. We consider that this can only be achieved through: (1) open access to scientific methods and data utilised; (2) open access to (analytical) code and methods; (3) open review of reported analyses/research. The proper application of such practices will reduce the innovation cycles in ATM, which is much needed by industry and society. Steps for forming an Open Science Alliance for ATM are described. We propose further initial, specific recommendations for supporting open data and improved access for research.

1 Introduction

The European Commission (EC), through the Horizon Europe programme, strives for open science, which is now a mandatory part of the funding actions: "Open science is an approach to research based on open cooperative work - that emphasizes the sharing of knowledge, results and tools as early and widely as possible. It is mandatory under Horizon Europe, and it operates on the principle of being 'as open as possible, as closed as necessary'" (European Research Executive Agency, n.d.).The open science is based on the FAIR principles: findable, accessible, interoperable, and reusable. It provides a wide range of benefits, intended to lead to:

- efficient academic research due to having more access to data, methods, results and publications;
- knowledge transfer and innovation;
- greater impact, visibility, and citation rates; and,
- eventually providing increased transparency and promoting higher levels of trust between research/governmental institutions and citizens.

Open science practices involve: open access (to publications), open data (sharing research data), open source/code (sharing software), open methodology (sharing models, methods, etc.), open peer review and open educational resources.

When these practices are suitably implemented, they foster greater efficiency in research through increased collaboration, higher levels of verification/validation, and reduced duplication. Furthermore, collaborations broaden the user community, improve methods and code testing, and enable more reproducible research (National Academies of Sciences, Engineering, and Medicine, 2018). Reproducibility is an important principle of the research process, as we can have more confidence in results that are confirmed through multiple findings, increasing the transparency of research. The ability of interested and independent researchers and practitioners to validate findings requires a spectrum of open science practices, including open access to the underlying data and analytical code.

Each field of science has slightly different characteristics and requirements when it comes to research and open science. Here, we are focusing on research performed within the Single European Sky Air Traffic Management (ATM) Research (SESAR) project. SESAR research is based on public, EC funding[1] and, as such, it should be, and is generally freely available to the research community and industrial or institutional stakeholders.

In general, public deliverables of the funded projects are available through the EC's CORDIS platform (European Comission, CORDIS, 2024), and many of the research deliverables from the first SESAR programme can be found in the EngageWiki (Engage KTN, 2022). Note that not all of the research, especially of the higher (technological) maturity, performed in the past years, is public.

Research papers from SESAR exploratory research (ER; i.e. research mostly at low technology readiness levels (TRLs[2])), are published through the SESAR Innovation Days conference website (SESAR Joint Undertaking, 2024), and other related (peer-reviewed) ATM/aviation conferences, and through open access articles in academic journals (as requested by Horizon Europe rules). The more mature research, the results of SESAR industrial research (IR), is published in the form of

[1] The research programme is managed and executed through the SESAR 3 Joint Undertaking, previously the SESAR Joint Undertaking.
[2] TRLs represent a method for estimating the maturity of technologies, initially developed by NASA.

SESAR Solutions[3] (SESAR 3 JU, 2023), which contain the specifications to enable the industrialisation of the Solutions.

Even though research *results* are generally accessible, over the years of running the SESAR programme, *requests for open data* access became increasingly articulated. In particular, this goes hand-in-hand with the emergence of data-driven methods, especially artificial intelligence (AI) / machine learning, which require much larger (e.g. longer-term) datasets.

However, limitations exist in sharing underlying datasets across different SESAR projects, or making these datasets available to other researchers and practitioners. A growing number of publishers encourage authors to share the underlying data or provide access to these data (e.g. Elsevier[4], Springer[5]). In various fields this practice is already a requirement. Within air transport and air navigation oriented research, the Journal of Open Aviation Science (JOAS) builds on ensuring reproducible research by providing access to the articles and related datasets. In this paper we focus on open data: the issues surrounding access to data and the benefits of opening this. Furthermore, we propose a roadmap towards an Open Science Alliance for ATM, a community dedicated to practising open science, for which open and accessible data is the foundation.

2 Background

The majority of ATM research in Europe centres on the SESAR project, which was established by European Council Regulation No. 219/2007 (Council of the European Union, 2007) with the ambition "to integrate and coordinate research and development activities which were previously undertaken in a dispersed and uncoordinated manner". The development and validation of SESAR Solutions is managed by the SESAR 3 Joint Undertaking (JU), which was first instituted in 2007, and extended in 2014 (Council of European Union, 2014), and in 2021 (Council of the European Union, 2021). The SESAR 3 JU is built on a consortium of industry and research organisations.

The SESAR programme covers different aspects of ATM, incorporating new topics, such as: improving the environmental impact of ATM; enabling multimodality and new types of air vehicles (e.g., drones, unmanned aircraft systems, air taxis, super-high altitude aircraft, hydrogen and electrically propelled aircraft); integrating new (for ATM) techniques and technologies (e.g., machine learning, AI).

The SESAR innovation pipeline flows through three phases, see Figure 1:

- **Exploratory research (ER)** that looks into new concepts, emerging technologies and methods, covering technology readiness levels (TRLs) 0-2. The attained knowledge can be transferred to the next phase/higher TRL level.
- **Industrial research (IR)** assesses, validates and verifies various technical and operational concepts across several performance areas. A concept brought to TRL6 is considered a SESAR Solution[6].
- **Digital Sky Demonstrators** which test the chosen SESAR Solutions in real life operations, and encourage the early take-up of Solutions. Some demonstrations collect needed evidence to mature the Solutions to TRL7 or even TRL8.
- **Fast-track innovation and uptake** activities may start at low TRL, like TRL2, aiming to deliver new products and services to the market, at TRL7.

[3] A Solution represents a change in the way ATM is performed. ATM Solutions are new operational concepts, procedures and relevant technologies.
[4] https://www.elsevier.com/about/policies-and-standards/research-data
[5] https://www.springernature.com/gp/authors/research-data/research-data-publishing
[6] Note that SESAR 3 JU can fund research only up to TRL6 (Ravenhill, et al., 2017).

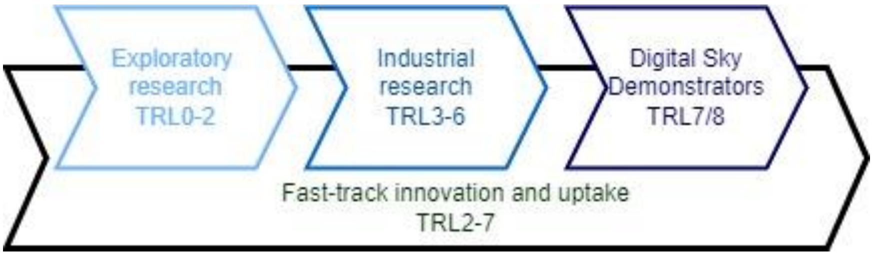


Figure 1. SESAR innovation pipeline, adapted from (SESAR Joint Undertaking, 2023).

SESAR funding is administered through open calls for proposals, where proposals are evaluated by experts, and the best ones are funded. The calls are separated by phase, where initial phases feed subsequent ones. This means that for example, the research matured to TRL2 in Exploratory research, would be included in a call for Industrial research. There are no guarantees that the same consortium of researchers would submit a proposal for the Industrial research, or that the submitted proposal would be retained for funding. See Table 1 for call overview across phases.

Table 1. SESAR pipeline, funding and datapack requirements.

	ER	IR and Fast-track	DSD
TRL	0-2	3-6	7,8
Funding through open calls	100%	70%	70%
Budget	1-2 million	4-15 million	10-35 million
Datapack requirements[7]	Feasibility	Prototyping, simulation, shadow mode, operational trials to complete: performance assessment, CBA, safety, security, human factors and environmental cases	Operational trials to complete: performance assessment, CBA, safety, security, human factors and environmental cases

Furthermore, the Covid-19 impact "shows that we need an ATM system that is sustainable, scalable, and resilient, which does call for the envisioned transformation" (Bolić & Ravenhill, 2021) (SESAR Joint Undertaking, 2020) and faster innovation cycles. This is also a requirement recognised by the international community. For example ICAO's Global Air Navigation Plan[8] identifies higher levels of agility and flexibility as key features of the future air navigation system. The long development cycles (for example, it took almost three decades for URET, an ATM decision-support tool to be deployed (Bolic, 2012)) prohibit the full exploitation of technological advances, further amplified by regulatory approach to system certification based on safety. The Wise Persons Group report (Wise Persons Group, 2019) identifies the slow technology uptake and ineffective regulatory processes as issues (among others) to be tackled for more efficient ATM. When a funding environment is also put into the mix, open science principles become the main enabler of faster development, permitting full *reproducibility* through access to results, data and methods of completed research projects. For example, if low TRL level results cannot be fully

[7] SESAR Solution datapacks contain safety and performance requirements, interoperability requirements operational service and environment description, technical specifications and interface requirements, and cost-benefit analysis (CBA).
[8] https://www4.icao.int/ganpportal

reproduced by the consortia (or other interested stakeholders) aiming to mature them to higher TRL, the development time will likely be longer rather than shorter.

Reproducibility can be represented as a spectrum (Marwick, et al., 2017), (Spinielli & Koelle, 2019). On the basis of this spectrum (Figure 2), the desired goal is to provide the ability to verify and validate research findings by being able to recompute the analytical results on the basis of a given dataset and insight into the approach/methodology and workflow of the analysis. This allows for the reproduction of consistent results. In Figure 2, the publication of a paper, including the associated results, represents the lowest level of reproducibility (in fact it is not independently reproducible at all) and as such, to be provocative, can be considered as scientific advertising activity. High reproducibility accordingly requires the sharing of the code and data underpinning the research/findings, a proper description of the methodology, and final paper/report productions. The state of the art in terms of reproducibility in ATM lies at the low end of the spectrum (in the majority of cases, just publication), mainly due to data sharing constraints. Non-reproducible research results hamper the efforts of further maturing the same, as it often requires starting from scratch when setting up the methodology, validation and verification, thus prolonging the time of (an ATM Solution) development. Several on-going developments postulated in this paper (see Section 5) are aimed at improving this situation and form the building blocks of an Open Science Alliance.

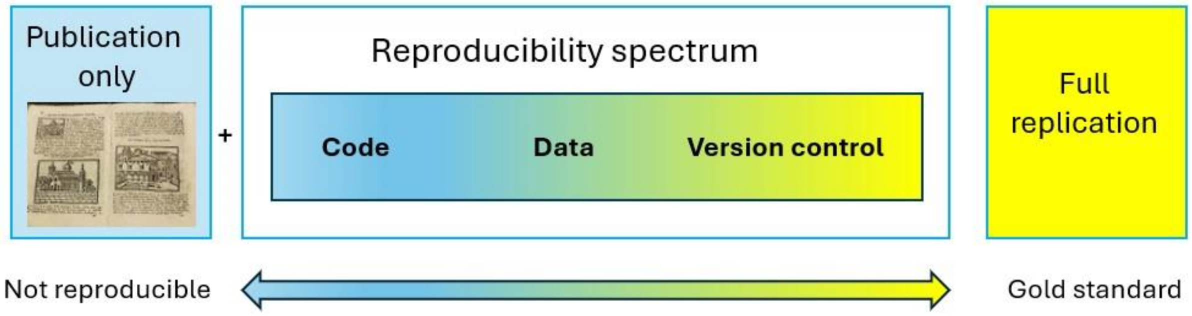


Figure 2. Reproducible research - reproducibility spectrum, adapted from Fig. 2 in (Marwick, et al., 2017).

During the SESAR2020 programme, the Engage Knowledge Transfer Network (KTN) with the goal to promote and facilitate the development of air traffic management research in Europe was instituted. The KTN ran from January 2018 to June 2022, and now continues as Engage 2 (Engage 2, 2023), launched in June 2023. The KTN was tasked with promoting and *facilitating* the development of ATM research in Europe, through the development of multiple, parallel activities. One of these was the EngageWiki (Engage KTN, 2022), which collected most of the research outputs since SESAR1, clustered by topic similarity (using semantic similarity analysis) in an interactive research map of ATM (see Figure 3). This opening of the access to European ATM research also served to inform the ATM future concepts roadmap (see Figure 4, which shows how previous (SESAR) research connects with the ATM Strategic Research and Innovation Agenda. The last stage in the figure, the Horizon flagship activities, represents the future concepts area of the roadmap, i.e., the examples of topics addressed in less detail in the SESAR past and planned programmes. A pool of previous (open) research results and various future roadmap documents was analysed and used to inform future research needs. This roadmap is being further developed under Engage 2. The interactive research map and the ATM future concepts map from the EngageWiki are just two examples that showcase what can be gained from open science principles.

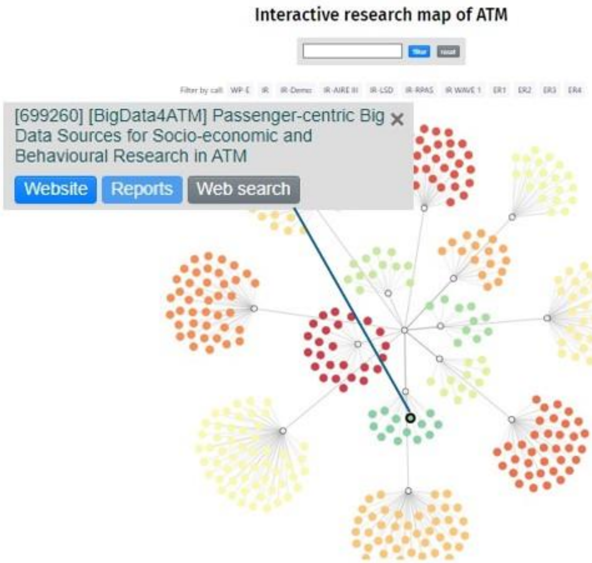


Figure 3. *EngageWiki research clusters.*

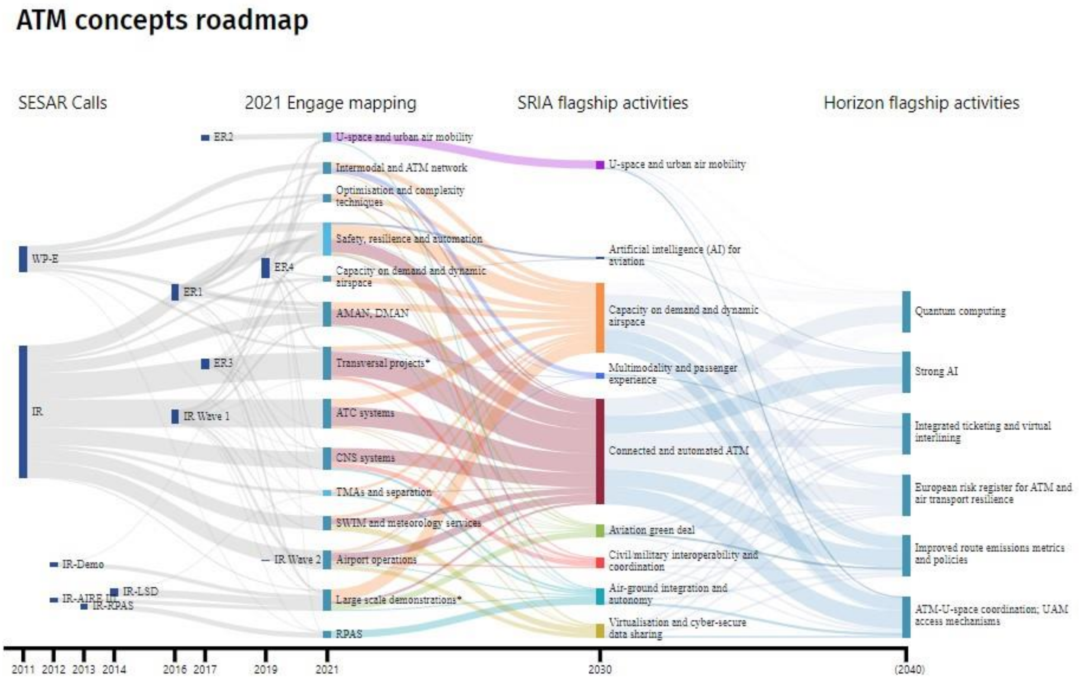


Figure 4. *EngageWiki ATM concepts roadmap.*

More importantly for the Open Science Alliance, the KTN consolidated views on the most important research barriers and enablers, with the associated lessons learned (Cook, et al., 2022). One of the most cited issues was data availability, which we term *open data*: in particular, we focus on its accessibility. This has been recognised by various stakeholders as a bottleneck in SESAR's ER, as vital data are often difficult to acquire, and often subject to non-disclosure agreements and/or other limitations, such as restricted use for one specific project. As such, this lack of open data, presents a "barrier to improving experimental comparability across projects" (Cook, et al., 2022) and precludes reproducibility. The KTN collected opinions of many projects, PhD students and workshop participants, where a common denominator was that they spent approximately 6 to 12 months (sometimes more) in acquiring, consolidating and cleaning data. The data required for

ATM research is composed of different types, and the exact requirements vary across the projects. Some data can be obtained freely (e.g. EUROCONTROL R&D data, ADS-B data from the OpenSky Network), some need to be paid for (e.g. schedule data, passenger itineraries and fares), and some need to be acquired from multiple sources if a greater geographical area is being researched (e.g. MET lightning or radar observations, public transit schedules) (Cook, et al., 2022). Since, frequently, some sort of licensing and non-disclosure agreement is required, the input data cannot be shared. The extent to which results can be shared is often dependent on the specificities of the licensing/agreements. However, if the input data cannot be similarly shared, achieving the desired comparability and reproducibility across projects and between research groups becomes close to impossible. The continuation of the Engage KTN, as Engage 2, furthers the open science principles through a set of actions. The founding and promotion of an Open Science Alliance are among them, which will be discussed in Section 5.

To sum up, a faster innovation cycle (i.e., less time in the innovation pipeline) in ATM requires reproducibility that is currently lacking. The entire ATM innovation pipeline is **dependent on data**, for research, development, industrialisation and even after deployment, for post-ops assessments. The easier it is to access necessary data, the shorter the innovation cycle becomes. To shed light on the reasons for such a state of affairs, we first describe the availability of data in ATM (in general terms), followed by an illustration of barriers, and then outline the opportunities that access to open data and adoption of open science principles could bring to ATM.

3 General availability of data for ATM research

Here we describe the availability of some of the most frequently used datasets in ATM research, and the impacts on such research. The list and usages are not exhaustive, but serve as an illustration of the ever growing need for open data.

ATM manages traffic and as such, almost all research needs some form of **flight data**. EUROCONTROL[9] provides free access, under licence, to archive data for R&D[10] that is two years old and limited to 4 months (i.e., March, June, September and December) per year.

Whilst it is often best practice to use the most recent data available in research, sometimes we specifically wish to use older data (e.g. to represent a period before a crisis, or to align with the availability of other data). Two-year old data are not always most appreciated by scientific publications, and it has been suggested (Ketchen, et al., 2023) that they should be used only in specific cases, such as when "examining a past event" or when "recent data are not available". Older data should especially be avoided in areas where it may compromise or bias the subject-matter under investigation, which needs to carefully considered in ATM. The time lag can also cause other issues, such as delays to post-Covid data release, where researchers await more 'normal' (recovered) traffic data.

These EUROCONTROL R&D flight data cover commercial flights (no military, state and general aviation) trajectories, last-filed and actual flight plans. The trajectories contain crossings into flight information regions (FIRs) and ATC unit airspace. The data are certainly useful for research into strategic (i.e., from planning phase) and some tactical (i.e., actual operations) innovations in ATM that deal with ATM performance -- airborne and at airports. For example, as the data contain aircraft registration numbers, one can investigate reactionary delays and turnarounds.

---

[9] EUROCONTROL is a pan-European, civil-military organisation dedicated to supporting European aviation. It is also a Network Manager for European air traffic network.
[10] https://www.eurocontrol.int/dashboard/rnd-data-archive

However, the utility of the data is often impaired by restricting it to four months, when considering the nuances of seasonal variability. This may mask changes to the network, connectivity patterns, operational concepts, weather or other short-term incidents.

Moreover, these data do not contain any information on air traffic flow management (ATFM) actions and/or consequences. When an air navigation service provider (ANSP) deems that in the coming hours their airspace (or portions thereof) is likely to be overloaded by the planned traffic, an ATFM regulation is usually activated. The regulations restrict the traffic by assigning delay to the flights (at the departure airport) planned to cross the said airspace. The management of ATFM actions, or, in other words, demand-capacity balancing, is an important area of research and post-operations analyses in European ATM. Research on demand-capacity balancing is thus rather restricted if one has access only to the EUROCONTROL R&D data.

Further, originally-filed flight plans are not included, thus the original intention of the airspace user is not known, nor is it possible to (accurately) model the situation *before* responses to flow-management regulations (see (Dalmau, 2024) for a discussion of differences between the R&D and operational datasets when assessing the reactionary delay, for example). The trajectories here are not exactly 'true' trajectories, but approximations between the filed and actual trajectories, as they diverge only when the difference between the two is large enough[11]. Thus, such data are not suitable for more precise analyses, such as those on operations around airports.

EUROCONTROL's Network Manager maintains another, much more expansive database, the Demand Data Repository (DDR2), which is the source of R&D data (i.e., a subset of it). DDR2 contains the last-filed flight plans, regulated plans, actual trajectories, and details on the ATFM regulation delays assigned to flights. Further functionalities include forecasting and visualisation tools. However, due to the high costs, both in monetary and data storage and management terms, DDR2 is available mainly to operational stakeholders: ANSPs, airports, airlines, and other airspace management actors, who use the data for post-operations analyses, checking the envisioned changes in their domain, simulations and the research projects in which these stakeholders participate.

More precise, actually flown trajectories are available from the OpenSky Network. The ADS-B/Mode S messages broadcast by aircraft contain the positioning, altitude, timestamp and other information. The frequency of certain messages can be quite high (e.g., twice per second for ADS-B) enabling creation of precise trajectories, at the expense of high memory and data preparation requirements. Free access to OpenSky Network's historical data (for research purposes), combined with shared tools make data acquisition and preparation faster (while still not insignificant), shortening the preparatory phase of research projects. Moreover, the significant amount of data available is conducive to the application of data science, machine learning and artificial intelligence techniques, that require large sets of data for model training.

Airspace and airport capacity data. Airspace data[12] refers to the division of airspace into sectors. Each air navigation service provider (ANSP) divides its airspace into sectors, portions of airspace controlled by a controller team. Depending on the traffic, the supervisor decides which opening scheme (i.e., which sectors) to activate, to provide the best service for the foreseen traffic, given the staffing levels available. The airspace data contains the indications of declared capacities of sectors, under different operational conditions, and the declared airport capacities. Currently, the airspace data is not available for academic research as it is not available in the R&D archive. Even though air traffic still has not reached the 2019 levels[13], delays have reached the levels of the capacity crisis in 2018/2019 (Performance Review Commission, 2023), pointing to the need for further

[11] For example, 5NM lateral difference or 10 minute delay.
[12] Also referred to as environmental data -- the environment where operations take place.
[13] https://www.eurocontrol.int/Economics/DailyTrafficVariation-States.html

developments in demand-capacity balancing. This requires understanding the airspace and airport environments, their capacity and traffic characteristics.

Investigating the impact of delay on the network and, more importantly, on passengers, requires information on **schedules and fare data**. In Europe, schedule data are available to purchase from providers such as Cirium or OAG Aviation. Airline passenger and fare data are available from global distribution system providers and IATA, for example, and are often extremely expensive. Airport passenger data are available at the airport level from a variety of sources, including civil aviation authorities and individual airports (usually for purchase), or freely from Eurostat (Eurostat, n.d.) at more aggregated levels. Commercial providers impose limitations on data usage: not sharing the purchased data is a standard requirement (understandably, from the business point of view), but this can also extend to the usage of such data (which can sometimes be limited to the research project for which the buyer enters into contract).

In the USA, the Bureau of Transportation Statistics (BTS) makes scheduled times available for the 17 airlines that "have at least 0.5 percent of total domestic scheduled-service passenger revenues" (Bureau of Transportation Statistics, n.d.), where this reporting is required by Federal Aviation Administration regulations. This has led to a significant volume of European research (including PhDs) using these datasets, when schedule information is needed. The aviation data available through the BTS website include various economic metrics relating to passengers (e.g., number of passengers carried, per market), fares (e.g., average fare from a set of airports, average fare on about 1000 domestic routes, specifying the largest and smallest airline in terms of market share, etc.), operating revenues (divided by market), load factors, available seat-miles, cargo carried, fuel cost and consumption, etc. In Europe, Eurostat database contains some statistics on the number of passengers, seat-miles and cargo carried, but these are aggregated either by airport or state. No data on fares, operating revenues or fuel are available.

Recent years have seen the introduction of **multimodality** topics into ATM research, as the EU is striving for improved mobility. To address multimodality challenges, additional data are needed from other modes of transport (especially rail, e.g., as an appropriate aviation short-haul substitute), which is often even harder to obtain than aviation data. The Digital Europe Programme[14] that focuses on bringing digital technology to the EU, is funding project aimed at creating European mobility data spaces[15]. The initial goal is to map existing mobility data-related ecosystems and initiatives, and to propose a common European mobility data space. These initiatives demonstrate that mobility data are recognised as important, but also that they are not easy to find and access, for both academia and business. As an example, train schedule (historic) data is not usually freely available (it differs across states), but may be obtained from rail associations. Train capacity and occupancy data are often collected only at the single company level.

Weather impacts are different across the flight phases (e.g., take-off, climb, cruise, etc.), for which **meteorological or MET** data are needed. The provision of MET data is available[16] for aviation purposes, it is standardised and regulated (ICAO, 2018) (European Commission, Directorate-General for Mobility and Transport, 2017), and in principle freely available for research purposes. These provisions cover certain types of observations and reports, such as Meteorological Aerodrome Report (METAR) and Terminal Area Forecast (TAF), and these ('live' ones) can be easily accessed by any interested party. Problems can arise when historical data are needed, as not every provider of this information (usually each state covers its own territory) can provide archiving services. An example of a collaborative platform that collects and shares data is Iowa

---

[14] https://digital-strategy.ec.europa.eu/en/activities/digital-programme
[15] E.g., https://mobilitydataspace-csa.eu/
[16] In Europe, the provision of MET data is financed through the route charges paid by flights using the air navigation services, MET being a part of it.

Environmental Mesonet (which collects environmental data) where METAR and TAF historical information can be found[17]. The European Centre for Medium-Range Weather Forecasts[18] provides different products for research and public access, such as the ERA5[19] atmospheric reanalysis of the global climate, and many others are available from the Copernicus programme. These datasets contain vast amounts of atmospheric data, but might be cumbersome to prepare and analyse for transportation purposes, as these are not standard transportation dataset. ERA5 reanalysis data is gaining higher visibility in air transport and ATM research as it provides data on atmospheric conditions based on the combination of model and meteorological observations around the globe. This is essential to address emission impact along the trajectory of a flight as METAR/TAF data only allow to derive weather conditions in close proximity to an aerodrome.

For many research projects there is a need to model aircraft behaviour in terms of acceleration, climb or descent performance, including the identification of optimum operating levels. This requires access to type-dependent and design-specific aircraft performance data. For **aircraft performance data** - EUROCONTROL's BADA4 data has become an industry standard. The downside (for expediting the research process) is that it can take a long time to get access approval, which is given per project only. A further constraint of the BADA licence agreement is that it excludes aircraft type-by-type comparisons. The availability of BADA and its use-constraints lead to work on an open source aircraft performance model (openAP, (Sun, et al., 2020)).

Most ATM problems require multiple data sources. Sourcing and preparation of data from any single database requires time, which can grow significantly if information from different datasets is needed. As a first step, the sharing of the underlying data can help other researchers interested in a similar analysis. For example, the underlying dataset of arrival trajectories for a paper 'Environmental inefficiencies for arrival flights at European airports' (Olive, et al., 2023) is available at https://data.4tu.nl/articles/_/20411868. This may reduce the effort to download and clean the trajectory data from its source. It is also the basis for reproducible research (cf. above) as other researchers may independently verify and validate the results on the same analytical data.

4 Barriers and opportunities

All ATM research projects need data. As ATM systems are complex, interconnected and heavy data consumers and producers, anything innovative also needs to be based on appropriate data, often from multiple sources. Within this context, it often takes between 6 and 12 months to obtain such data for research purposes. Taking into account that the duration of most of the SESAR (and Horizon Europe) projects is 24 months, and that modelling, testing, and validation cannot be performed without data that may arrive only half-way through the project duration, the possibilities of ultimate progress within a single research project are often restricted by data availability. Furthermore, most of the data in ATM is treated as confidential by the data providers, which, in turn, requires various licensing or non-disclosure agreements. The impact of this is twofold -- the time required to complete the appropriate paperwork and, more often than not, the limitations on what the data can be used for: very frequently linked to a specific project only, thus precluding any input data sharing and dictating the shape or format of results' sharing.

4.1 Barriers

The introduction highlighted the benefits of an open science (data) approach. An underlying requirement is the cultural change embracing transparency, reproducibility, and a public

---

[17] https://mesonet.agron.iastate.edu/request/download.phtml
[18] http://www.ecmwf.int/
[19] European Centre for Medium-range Weather Forecast Reanalysis v5.

discourse. This change will require removing barriers to opening data. Such barriers can be broadly categorised as (see Table 2 below for an overview):

- reputational repercussions;
- privacy concerns (both at individual and organisational levels);
- business considerations (comprising, *inter alia*, the confidentiality of business-critical data, intellectual property protection, legal liability or restrictions);
- national security and defence issues;
- costs (associated both with the collection, processing, and storage of 'massive' data in general, and the provision/maintenance of open data).

A recurrent theme when discussing open data is **data ownership** (and, implicitly, the associated authority controlling its use). Air transport is a public transportation domain. As a public service, involving various stakeholders and organisations, their processes and operations are governed by a variety of rules and laws. Across the globe, states have implemented laws that address any interference with the signal transmission between airborne and ground systems. This is targeted to ensure the safe execution of flight and -- in principle -- prohibits non-aviation stakeholder from interacting with the data or transmissions. Traditionally, air transportation data collection, processing, and distribution was dependent on dedicated technologies (e.g., surveillance data required radar installations). From a pure research perspective, the costs of installing such technologies is prohibitive and poses technological complexities. In consequence, state organisations or service providers involved in air traffic control and management, including MET or airline operations are considered 'data owners' as it describes the entities and systems which predominantly had access and/or collected the data. Thus, historically, the technological processing requirements limited the wider public use and access to air transportation related data. In Europe today, majority of flight-related data is integrated and stored by EUROCONTROL, in its role of the European Network Manager. The data is constantly used for planning (i.e., network and flight planning (EUROCONTROL, 2023)), operationally (e.g., when capacity regulations are needed) and post-operationally (e.g., bases for route charges, or assessment of performance plans (European Commission, 2019)) by the member stakeholders (ANSPs, airlines, airports). A good portion of the data usage is in service of performing tasks mandated by regulations. For example, the ANSPs are funded through the route charges, which are based on the actually flown trajectories (European Commission, 2019). However, calls for opening the data for research activities often stumble against the concerns around reputation (e.g., fear of comparisons, especially if perceived to be unfair) and privacy (e.g., tracking a specific person).

In light of the emerging debate on climate change -- including the role and impact of air transportation -- a wider interpretation constrains the exclusive private authority over data ownership. The continual pressure of climate impact will likely tailor exemptions related to privacy or restricted access to public transport data: individual rights to support broader societal well-being. In this context, access to data of public interest gains a higher momentum. For example, the Freedom of Information Act in the United States provides an instrument to request access to -- publicly funded -- records and data, including air transport. Accordingly, a prerequisite for open science/data is a proper separation of the legitimate use of data from the concerns of individuals, organisations, and legal/regulatory regimes.

While of a different nature, (concerns regarding) loss of reputation, confidentiality, and national security/defence fall into a similar category. The concern is the disclosure of information, or behaviour, or level of capability, that may result in 'unfair' assessment/comparison or leaked secrets. Similar to operating a vehicle on the streets, the operation of aircraft follows certain rules and requires, *inter alia*, the display/dissemination of unique flight identification, holding of insurance, demonstrating the technical reliability of the vehicle, to ensure safe operations in the public space, etc. It follows that the identification of aircraft -- under the assumption of a legitimate public interest -- is not a privacy or secrecy issue and subject to non-disclosure.

Table 2. Barriers to opening data.

Barrier	Type	Importance
Data ownership and control	Legal/regulatory	Issues around who owns and controls data usage, especially when data is collected from multiple stakeholders, even if at public transportation sectors
Reputational repercussions	Reputational	Risk of negative perception or damage to reputation from data analyses
Privacy concerns	Privacy	Concerns about individual or organisational privacy being compromised
Business considerations	Business	Issues related to confidentiality, legal liabilities, and business-critical data
National security and defence issues	National security	Concerns over the disclosure of sensitive information that could affect national security or defence operations, needing to define what is legitimate usage of which parts of data
Costs	Financial	Expenses associated with data collection, processing, storage, and maintenance. Prohibitive costs and complexities in installing necessary technologies for data collection and processing
Public interest and societal well-being	Ethical/public policy	Balancing individual privacy rights with the broader societal benefits, especially in the context of climate change and mobility

Particularly, the debate about the latter received public attention over recent years. Investigative journalism or the use of such information by environmental interest groups or enthusiasts led to a series of public arguments. The following lists a few examples of these episodes during which privacy concerns were raised:

- the CEO of the luxury goods brand LVHM, Bernard Arnault, sold his private jet after reports emerged about the frequency and destination of his trips (Insider, n.d.). To shield movements, the company now leases aircraft that are not directly linked to the CEO/company;
- the entrepreneur Elon Musk (Tesla, X) drew ire with the ElonJet, an initiative that tracks the movements of Musk's jet. Musk argued at one point in time that the openly available information on his location may also endanger the safety of his family or him. On a similar note, the singer Taylor Swift triggered a cease-and-desist order for the tracking of her private jet movements;
- the coverage of the flight planning system disruption of UK's ANSP, NATS (James Haydon, 2023) is an example which shows how publicly available data can be used to present the combination of unfortunate system processes and undetected implementation issues that could lead to reputational damage and possibly higher scrutiny by regulatory bodies.

While celebrities generally enjoy the same privacy rights as citizens, it must be noted that the desire for higher privacy does not expand into the aforementioned public transport domain, in particular if the data are post-operational. On the other hand, national security and defence operations may enjoy higher privileges and specific rules. These may ultimately entail limiting the identification of aircraft or their interaction with other sensors. It is also important to differentiate the rules under which such operations are performed. Under the ICAO framework, the sovereignty of a state and its services are undisputed. Accordingly, procedures are in place to shield the identification of state-operated aircraft in such circumstances. The principle of open data *per se* does not infringe these considerations, as it relates to the operation of flights within the existing rules. The inherent detection of national capabilities, response times, or mission profiles can therefore be subject to restrictions of the use of data.

The emergence of 'big data' and recent data processing and storage techniques are not cost-free. Next to the pure technological operating cost (e.g., server hardware infrastructure, electricity), the collection of the data, developing data interfaces, converting received data into a format to be stored, and potential data cleaning and integrity checking present additional costs and efforts (Schäfer, et al., 2018) (Strohmeier, et al., 2015). There are synergies of scale, if a shared vision is supported by a number of individuals and organisation. Such an alliance of the willing / like-minded can be found in crowd-sourced or community-based approaches. This helps to share the burden, increase coverage, and strengthen the collection of data. Challenges of maintaining and sharing collected and (partly) pre-processed datasets, could be overcome through different established repositories, such as the OpenSky Network or Zenodo[20]. Web-based networks of enthusiasts, researchers, and institutions could help facilitate the effort of data collection. Further examples of overcoming cost/effort barrier can be found in the Open Aviation Data Github repository[21].

Aviation is a data-rich environment, and while certain operational data are readily processable, other datasets require a substantial suite of sensors. MET data, particularly on a wide-area level and high granularity are based on cost-intensive sensors (e.g. satellite imagery, air mass related sensor measurements, LIDARs). Particularly in these fields, public funding will be an essential enabler for open access data in the future. To date, ESA and NASA are subscribing to open science (National Academies of Sciences, Engineering, and Medicine, 2018) (European Space Agency, 2024) and are providing free access to a vast amount of MET and atmospheric information. Public funding was also used for establishing pre-processed climate change data and host these data in Climate Data Stores[22].

4.2 Opportunities

Today's environment is characterised by an increasing number of isolated developments that can serve as the foundation for the envisioned open science/data approach put forward in this paper. The last decade or so has seen a general increase in data literacy by individuals, modern data processing and analysis techniques, and a general increase in available datasets. At the political level, the regulatory players (e.g., the European Commission, national governments) have embraced the benefits of open data, and across Europe it is a common feature to couple public funding with open access to both results and underlying data for research, policy analysis, etc., on the reuse of data between different research projects (and potentially competing organisations).

However, the current practice is still in its infancy, as several exemptions exist surrounding the protection of intellectual property rights at the same time. Horizon Europe provide the "as open as possible, as closed as necessary" clause, allowing results and data kept "closed if making them public in open access is against the researcher's legitimate interests." (European Research Executive Agency, n.d.). The clause makes sense when the results of the project lead to direct industrial implementation, but one can argue not, otherwise.

Within the aviation domain, a continuously growing number of initiatives to crowd-collected and acquired data and sharing exist - at least within the respective community. This phenomenon is supported by web-based platforms for data storage and processing. These community-based efforts range from aircraft inventories, aeronautical information, to air transport movement data (Schäfer, et al., 2014) (Engage KTN, 2022) (OpenAIP, 2023). These interest groups may have different backgrounds (e.g. technology discussion groups, aviation enthusiasts, aircraft tracking,

[20] Zenodo is a general-purpose open research repository operated by CERN: https://zenodo.org/.
[21] https://atmdata.github.io/sources/
[22] For example, https://cds.climate.copernicus.eu/. For example, ERA5 represents the fifth generation of reanalysis data, a combination of measurements and modelled meteorological data (Hersbach, et al., 2020).

private flight planning, researchers). They share the same goal: to establish, complement, clean, and provide insights based on shared open datasets.

The political ambitions and goals surrounding the future of air transportation and the discussion of aviation's impact on climate change, call for higher levels of transparency, increased levels of operational efficiency as near-term measures, and a faster adoption of novel operational concepts. This requires implicitly a transparent monitoring of the changes and improvements in performance. Open data will make it easier to apply the performance assessment frameworks within research, such as SESAR, or operations, by ensuring the required transparency and reproducibility. Particularly within the (European) research domain this may enable a higher research ranking on the global stage, as traceability of (European) solutions and a comparison of such solutions in different operational environments are supported.

More broadly, open data goes hand-in-hand with open code and open publications, which we next address.

OpenSky Network, an example of open data community achievements

The OpenSky Network functions as a collaborative sensor network, collecting surveillance data for air traffic control purposes. Its core mission is to grant the general public access to real-world air traffic control (ATC) data and *promote the advancement of ATC technologies and processes*. Since 2013, the network has continuously gathered air traffic surveillance data, distinguishing itself from commercial flight tracking networks such as Flightradar24, or FlightAware, by preserving the original Mode S replies in an extensive historical database accessible to researchers and scientists across various fields. Initially, the non-profit network began with eight sensors in Switzerland and Germany but has expanded to include over 6500 registered receivers worldwide. As of 2023, OpenSky's dataset spans over a decade of aircraft broadcast data.

While its initial focus was on ADS-B, it broadened its data range to include the complete Mode S downlink channel in March 2017. More recently, it integrated other technologies such as FLARM and VHF. The dataset comprises over 35 trillion Mode S replies, experiencing a peak influx of over 20 billion messages per day. The global data reception of the OpenSky Network relies on a crowd-sourced network of receivers, primarily comprised of enthusiasts, academics, and supporting institutions. Each sensor's coverage is limited by the range of antennas' line of sight, typically around 400–500 km for the best-performing antennas reaching the radio horizon. The organic growth areas of such a crowd-sourced network effectively represent densely populated and wealthier regions worldwide. Between 2018 and 2023, the network's global coverage (see Figure 5) reached saturation, with many new sensors mostly enhancing reception at lower altitudes in areas already covered in Europe, the US, and other developed countries. However, notable coverage expansions are still observed in the Middle East, South Asia, and New Zealand. Geographical regions such as deserts and oceans lack ground-based coverage due to physical constraints, prompting commercial ADS-B providers to partially rely on space-based ADS-B or ADS-C data (Woodrow, 2019). OpenSky, too, will add ADS-C data in the future.

OpenSky offers several interfaces to access the live and historical data (using different underlying technologies). The OpenSky Network developed a very active research community, from different, related research areas: machine learning, ATM, security, Earth science, etc. The OpenSky community has developed extensive tooling for easy data access, processing and visualisation, most notably *traffic* (Olive, 2019) and *pyModeS* (Sun, et al., 2019), thus easing access and initial data preparation. The growth of the OpenSky community is helped by the yearly symposia, aiming at strengthening the exchange of data, methods and code. For further details on the history, architecture, and use cases of OpenSky, please refer to (Schäfer, et al., 2014), (Strohmeier, 2020).

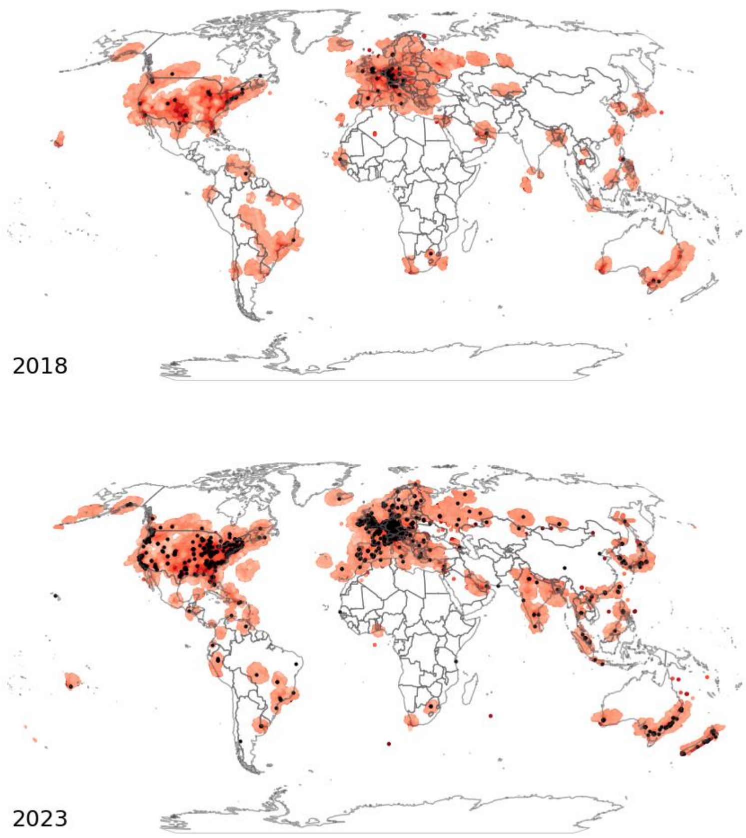


Figure 5. *OpenSky's global coverage in 2018 and 2023.*

OpenSky data has been used in more than 470 publications as of September 2023. While the trend is growing (see Figure 6), open data is clearly only used in a fraction of the air traffic data-based research conducted every year, as many other publications are using restricted-access data, for a variety of reasons, which are explained in the following section.

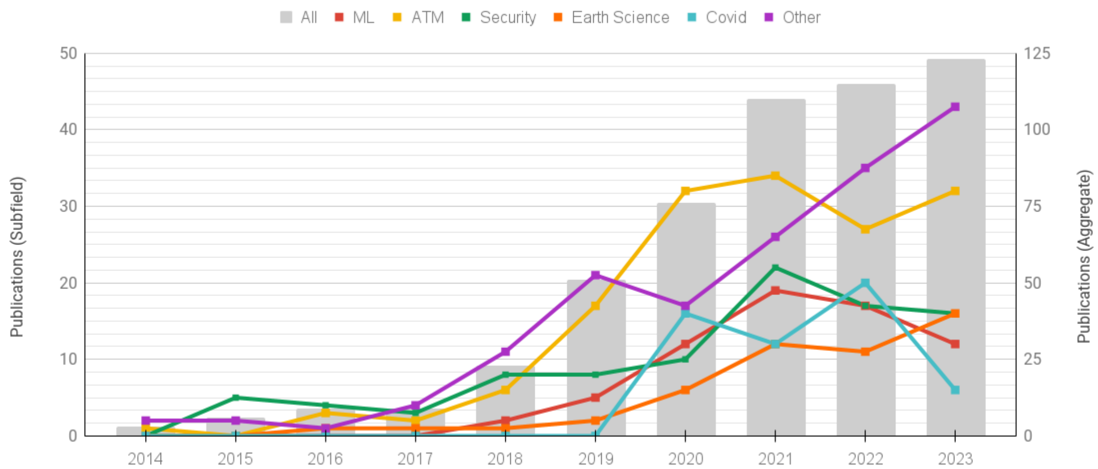


Figure 6. *Publications using OpenSky Network data from 2014 - 2023.*

Open performance data initiative

Ambitious political goals and strategic objectives ultimately require a 'democratisation' of access and use of ATM data. This would allow political decision-makers, strategic planners, practitioners and researchers to tap into a harmonised open data environment. As an initial step towards such an environment and level playing field, the Performance Review Commission (PRC) of EUROCONTROL launched the open performance data initiative (OPDI) (Performance Review Commission, EUROCONTROL, n.d.), aiming at closing the gap towards data accessibility and openness in European aviation research - supporting research validation and reproducibility.

The OPDI datasets are based on OpenSky Network crowd-sourced (ADS-B) data for the period 2019 to present. It comprises:

- A flight list, typically, flight identification (ID), Aerodrome of Departure (ADEP), Aerodrome of Destination (ADES), (UTC) day of flight, aircraft type,
- A list of milestones for each flight (see examples below),
- A 10-second state vector representation of the relevant flight trajectory as recorded by OSN. In a very simplified form, a state vector is a so-called point profile, i.e., a sequence of 3D positions and relevant timestamps.

Milestones are a way to summarise/synthetize a (typically 1-second samples) state vector for a flight to a set of "significant" events encoded as a 4D position (flight id, 3D + timestamp), milestone type and contextual info. Examples of milestones are:

- off-block (contextual info = parking position ID),
- runway (RWY) entry (contextual info = RWY ID),
- lift-off (contextual info = RWY ID),
- top of climb,
- Flight Information Region (FIR) crossing (contextual info = FIR ID1-FIR ID2),
- …
- top of descent,
- start of holding,
- end of holding,
- start of level segment,
- end of level segment,
- touchdown (contextual info = RWY ID),
- RWY exit (contextual info = RWY ID),
- on-block (contextual info = parking position ID).

The milestones dataset, for example, can be used to extract departure, arrival, internal, overflight statistics by using FIR crossing milestones, or in the estimation of CO_2 and NO_x emissions by modelling a flight via the milestones collected for the ground, departure, en-route and arrival data.

The availability of an open milestones dataset linked to a flight list and state vector flight profile is useful for research as a common reference upon which to build analysis and models. Milestone types are not fixed and new ones can be easily accommodated, allowing the dataset to grow with the needs of the research community. The work of cleanup, milestones extraction, and flight list identification is already available and can lift the burden of initial data acquisition and raw data preparation, which can require long lead times. Such datasets, freely available for research purposes make research reproducible, increasing research quality. The dataset could be used as a *de facto* standard for actually-flown trajectories. The OPDI dataset can also ease the project setup in allowing, for example, the pre-filtering of the subset of flights to consider, without having to start from a raw dataset.

Open code

Open code refers to sharing pieces of software and code (including functions) within a community, which may help with common data manipulation or calculational tasks, sometimes as libraries. Open code can be "openly accessed, inspected, modified, and enhanced by anyone" (Performance Review Commission, EUROCONTROL, n.d.). A good example can be found in the OpenSky Network community, which shares extensive tooling for easy data access, processing and visualisation. There are many such communities around Github, even one dedicated to ATM[23]. We mentioned transparency as a key feature, and open code is an essential building block of this. While open code may initially serve as a documentation and provides context, it also supports mutual learning and improvement. Other community members, researchers, or practitioners can inspect and improve the code, including runtime and resource efficiency, or improve the robustness by augmenting the code to catch fringe cases.

Open aviation journal

The Journal of Open Aviation Science (JOAS)[24] is a novel initiative started in 2022 and supported by TU Delft. JOAS is dedicated to advancing high-quality open science, open data, and reproducible scientific research within the aviation domain. Recognising a deficiency of such practices in current aviation research, the principal objectives are to foster reproducible research in aviation through the utilisation of open-source tools and open data, and to promote the replication of existing research within the aviation research domain. Submissions must include data, accompanied by publicly available code, and explanations, ensuring that the code is tested for reusability by other researchers.

5 Roadmap towards an Open Science Alliance for ATM

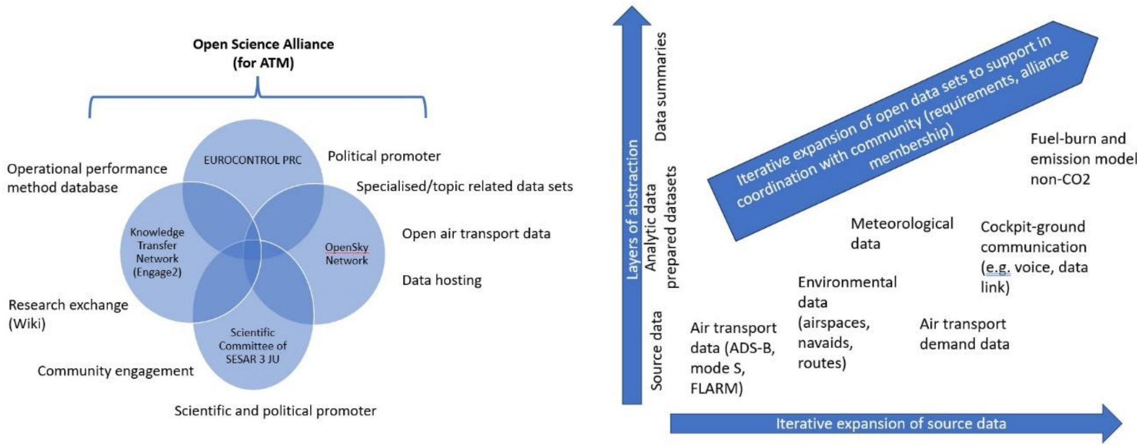
The preceding text describes open science, focusing on open data, the benefits it offers and where we are regarding open data in ATM research. It also describes some of the key challenges and opportunities. We now turn to proposing a roadmap towards an Open Science Alliance.

Mindful of the challenges and opportunities set out above, we propose the establishment of an Open Science Alliance, which fully encompasses the broader open principles described, not least open data. Such an alliance could build on the existing collaborations of Figure 7 (a), leveraging the existing progress and dialogue, and, moreover, commitment to the principles laid out therein. Similar to a jigsaw, joint initiatives and priorities would allow the expansion of the picture by adding more pieces. From an organisational level, alliance members may span the various fields of securing political support, methodological frameworks, supporting research and development, data collection / hosting / sharing services and platforms, and associated data feeders.

EUROCONTROL's PRC may promote further the values of opening data and offer their support in building the operational performance methods, and the definition of specialised datasets. The Scientific Committee (SC) of the SESAR 3 JU[25] (SESAR 3 Scientific Committee, 2023) identified open data as an issue requiring support and established the open data task force. This paper is a result of the task force's collaboration with the wider scientific community. The SC will continue its mission in identifying opportunities and promoting the need for open data.

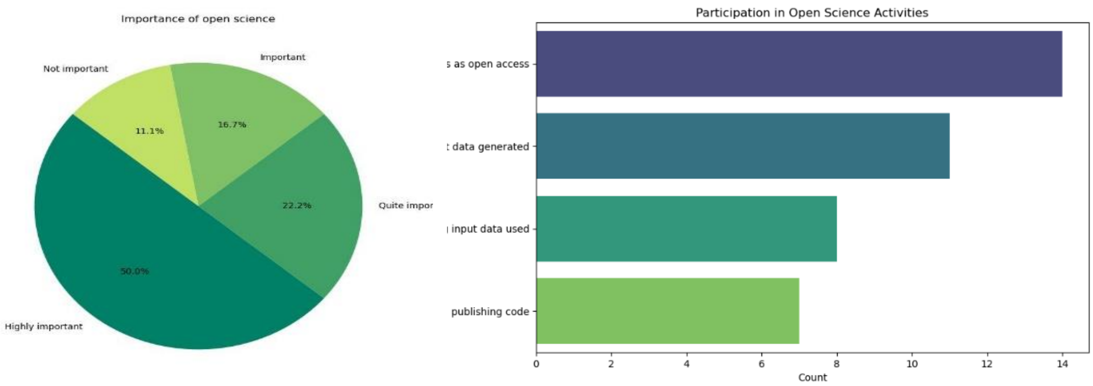
---

[23] https://atmdata.github.io/
[24] https://journals.open.tudelft.nl/joas
25 https://www.sesarju.eu/about-sesar/scientific-committee



(a) First step *(b) Possible expansion path*

Figure 7. Open science alliance for ATM.



(a) Importance of open science *(b)Open science activities practiced*

Figure 8. Engage 2 open science survey results.

The OPDI is a substantial first step, showcasing the alliance between EUROCONTROL's PRC and the OpenSky Network, which is already available online. The latter also has broadly and historically supported the collection and sharing of open air transport data. The further step consists of tight collaboration of the SC and Engage 2. Engage 2, as a SESAR KTN, is committed to open science principles, further development of the EngageWiki and, in particular, community building and outreach. An open data survey was launched by Engage 2, to try to determine the gap between industry and academia in ATM research, by considering the great importance of open science for supporting the intended growth, and offer the possibility to join the Open Science Alliance. Most of the respondents so far (it is still open), come from higher education or research institutions. Figure 8 (a), shows that the open science principles are important (highly important for around half of the respondents), and that, indeed, in ATM the most common open science activity is publication of articles as open access. All other activities are less frequent, publication of input data and code bringing up the rear. About 8/10 respondents expressed the desire to join the Open Science Alliance.

Further, the open science principles permeate all major actions supported by Engage: definition of thematic challenges in ATM research, support to PhD and MSc students in the ATM domain, and fast-tracking specific activities in support of the maturing ATM Solutions through cascade funding. At the time of writing, the winners of PhD funding through Engage 2 have been informed, and contracts are being drawn up. The next step is the school for the Engage 2 funded students (and all other interested students), scheduled for the beginning of 2025. In this first and subsequent schools, and the workshops planned around thematic challenges, open science principles will be promoted, and participants will be offered a possibility to join the Open Science Alliance.

To additionally promote open data, a key change to the EngageWiki is planned -- adding a collection of links to the published datasets (with the permanent link, e.g., publication on Zenodo), code and publications including existing summaries (thus also supporting 'first mile' accessibility for those wishing to avoid data extraction / preparation). Engage 2, the SC, EUROCONTROL's PRC and the OpenSky Network already collaborate on many areas of open science principles and will continue to do so, defining new initiatives in the future.

As identified in Figure 7 (a), and described in the text above, the initial, loose Open Science Alliance already exists. The initial **terms of reference (ToR)** of the Open Science Alliance would revolve around the following points:

- foster the Open Science Alliance by bringing together like-minded organisations and stakeholders to establish and expand the collection, preparation, sharing, and usage of open data / code / publications;
- promote and encourage the expansion of contributing members facilitating and opening internal processes and practices, and actively requiring the sharing and complementing the existing datasets of the alliance;
- when the alliance is better established, prioritise developments and priorities, and contribute by aligning organisational-level activities and contributions to the alliance's portfolio of activities, development areas, and the preprocessing / augmentation of datasets;
- maintain and continuously support the hosting and sharing services via the portal, and support the further development of the portal and technical improvements; and
- support the continual expansion of analytical datasets for sharing via (the) portal(s), both in terms of augmenting the existing datasets (e.g. increase temporal or geographical coverage), and also in terms of expanding datasets in line with the identified needs and priorities of the community / alliance.

Figure 7 (b) depicts a possible expansion path. Based on the core alliance identified above and the proposed initial ToR, further specific membership including executive bodies (such as the European Research Executive Agency) with the power to support the funding of specific recommendations (where required), and the wider scientific community and associations, further elaborating the requirements and ToR, would be expanded in an iterative process. The issue of governance could then be addressed a little further downstream, as the context and inclusive membership is more closely formulated.

Along with membership-level expansion, an alignment of the initiatives, programmes, and dedicated actions to support further development, alliance members could help to accelerate the availability of the respective source data, cleaned datasets, and pre-processed analytical data. Figure 7 (b) offers initial thoughts of additional datasets and the level of abstraction (i.e., layers of abstraction).

6 Conclusions and recommendations

We strongly believe in, and subscribe to, the practices of open science. Openness and reproducibility are two sides of the same coin and will ultimately enable trust and flexibility, and ensure the positive developments identified through this paper. Individual, organisational, and institutional sensitivities can range from privacy concerns to issues of liability. However, the benefit of adapting an open science approach will ultimately pay dividends. The challenges faced by air transport and ATM require higher levels of transparency. A fundamental enabler will be reproducibility - enabling independent verification and validation of reported impacts/results and achieved performance levels. We consider that this can only be achieved with:

- open access to the scientific methods and data utilised;
- open access to the analytical code and implementation of methods;
- open review of reported analyses/research.

This paper has identified an initial roadmap to establish an Open Science Alliance for ATM research in Europe. Such an open partnership would represent a stepping stone towards leveraging the benefits of open data, *inter alia*, to increase transparency and accountability, accelerate the validation and outreach of research, and promote faster innovation cycles. The cultural change required is to move away from considering data as the 'new gold', and make it an asset. Europe is in an excellent position to come together and realise such an alliance, which can be achieved by bringing together activities within different domains, through a concerted action on open science in the widest sense.

For specific consideration within the framework of this broad alliance, we further recommend, in no implied order of priority, to:

- together with regulators and policy makers, identify future data requirements as they pertain to the Green Deal and Digital Europe programmes, in the ATM domain (e.g., passenger itineraries and fares, schedules);
- support and finance a common licencing programme for commercial data required by multiple research projects to obviate the loss of finance and human effort of multiple spends on the same, or very similar, data (such as for airline schedules and passenger itineraries);
- review the requirements for domain-specific (e.g., environment, multimodality, MET, cybersecurity) data with the current SESAR projects, and collate these to be considered for common licences and/or for improved open access (e.g., by relaxing project-specific approvals for broader activities within SESAR ER and IR, such as for BADA 4 data) - this could be initially reviewed at the SESAR flagship level;
- consider the open science survey results, currently open by Engage 2, to specifically investigate the data needs and access challenges of the wider ATM research community.

Acknowledgements

The authors would like to thank all colleagues who contributed to shaping these thoughts over recent collaborations. Parts of the article originate from material from the Engage (1) KTN, Grant agreement No. 783287, funded by the SESAR Joint Undertaking. Parts of the article are covering ongoing efforts in Engage 2, funded by SESAR 3 Joint Undertaking under Horizon Europe, Grant agreement No. 101114648.

Data/Software Access Statement

Research data showcased in this publication are available from the following websites:

- https://www.opdi.aero/
- https://opensky-network.org/

Contributor Statement

Conceptualization: Bolić Tatjana, Koelle Rainer

Methodology: Bolić Tatjana, Koelle Rainer, Cook Andrew

Data Curation: , Goens Quinten, Spinielli EnricoStrohmeier Martin

Visualization: Bolić Tatjana, Strohmeier Martin, Koelle Rainer

Writing –Original Draft: Bolić Tatjana, Koelle Rainer

Writing –Review & Editing: Bolić Tatjana, Cook Andrew, Koelle Rainer, Spinielli Enrico, Goens Quinten, Strohmeier Martin

Conflict Of Interest (COI)

There are no conflicts of interest.

References

Bolic, T., 2012. Innovation Adoption and Adaptation in Air Traffic Control: Interaction of Organizations. *International Journal of Sociotechnology and Knowledge Development,* January, Volume 4, p. 17–29.

Bolić, T. & Ravenhill, P., 2021. SESAR: The Past, Present, and Future of European Air Traffic Management Research. *Engineering,* April, Volume 7, p. 448–451.

Bureau of Transportation Statistics, n.d. *Airline On-time Statistics.* [Online] Available at: https://www.transtats.bts.gov/ONTIME/ [Accessed 29 September 2023].

Cook, A. et al., 2022. *D3.10 Research and innovation insights,* s.l.: s.n.

Council of European Union, 2014. *Council regulation (EU) No 721/2014 of 16 June 2014 amending Regulation (EC) No 219/2007 on the establishment of a Joint Undertaking to develop the new generation European air traffic management system (SESAR) as regards the extension of the Joint Undertaking until 2024 (Text with EEA relevance).* s.l.:s.n.

Council of the European Union, 2007. *Council regulation (EC) No 219/2007 of 27 February 2007 on the establishment of a Joint Undertaking to develop the new generation European air traffic management system (SESAR).* s.l.:s.n.

Council of the European Union, 2021. *Council Regulation (EU) 2021/2085 of 19 November 2021 establishing the Joint Undertakings under Horizon Europe.* s.l.:s.n.

Dalmau, R., 2024. Probabilistic and explainable tree-based models for rotational reactionary flight delay prediction. *CEAS Aeronautical Journal,* July.

Engage 2, 2023. *Engage 2 KTN website.* [Online] Available at: https://engagektn.eu/ [Accessed 20 February 2024].

Engage KTN, 2022. *Engagewiki.* [Online] Available at: https://wikiengagektn.com/EngageWiki [Accessed 29 September 2024].

EUROCONTROL, 2023. *European Network Operations Plan 2023-2027.* s.l.:s.n.

European Comission, CORDIS, 2024. *CORDIS EU Research Results.* [Online] Available at: https://cordis.europa.eu/en [Accessed 16 February 2024].

European Commission, Directorate-General for Mobility and Transport, 2017. *Commission Implementing Regulation (EU) 2017/373 of 1 March 2017 laying down common requirements for providers of air traffic management/air navigation services and other air traffic management network functions and their oversight (Text with EEA relevance.).* s.l.:s.n.

European Commission, 2019. *COMMISSION IMPLEMENTING REGULATION (EU) 2019/317of 11 February 2019 laying down a performance and charging scheme in the single European sky and repealing Implementing Regulations (EU) No 390/2013 and (EU) No 391/2013.* s.l.:s.n.

European Research Executive Agency, n.d. *Open science.* [Online] Available at: https://rea.ec.europa.eu/open-science_en [Accessed 21 September 2023].

European Space Agency, 2024. *Open Science.* [Online] Available at: https://www.esa.int/About_Us/Digital_Agenda/Open_Science [Accessed 31 October 2024].

Eurostat, n.d. *Eurostat, The home of high-quality statistics and data on Europe.* [Online] Available at: https://ec.europa.eu/eurostat [Accessed 20 February 2024].

Hersbach, H. et al., 2020. The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society,* Volume 146, p. 1999–2049.

ICAO, 2018. *Annex 3 - Meteorological Service for International Air Navigation,* s.l.: s.n.

Insider, n.d. *The world's 2nd-richest man, Louis Vuitton's CEO, sold his private jet after people started tracking it on Twitter: 'No one can see where I go'.* [Online] Available at: https://www.businessinsider.com/louis-vuitton-bernard-arnault-sold-private-jet-people-tracking-twitter-2022-10 [Accessed 28 September 2023].

James Haydon, 2023. *UK air traffic control meltdown.* [Online] Available at: https://jameshaydon.github.io/nats-fail/ [Accessed 28 September 2023].

Ketchen, D. J., Roccapriore, A. Y. & Connelly, B. L., 2023. Using Old Data: When Is It Appropriate?. *Journal of Management,* June, Volume 49, p. 2541–2548.

Marwick, B. et al., 2017. Open science in archaeology. *SAA Archaeological Record,* Volume 17, p. 8–14.

National Academies of Sciences, Engineering, and Medicine, 2018. *Open Science by Design.* Washington(DC): National Academies Press.

Olive, X., 2019. Traffic, a toolbox for processing and analysing air traffic data. *Journal of Open Source Software,* Volume 4, p. 1518–1.

Olive, X., Sun, J., Basora, L. & Spinielli, E., 2023. Environmental inefficiencies for arrival flights at European airports. *Plos one,* Volume 18, p. e0287612.

OpenAIP, 2023. *Open AIP, free worldwide aviation database.* [Online] Available at: https://www.openaip.net/

Peng, R. D., 2011. Reproducible research in computational science. *Science,* Volume 334, p. 1226–1227.

Performance Review Commission, EUROCONTROL, n.d. *Open performance data - an enabler to further performance enhancements?.* [Online] Available at: https://www.eurocontrol.int/publication/performance-insight-5-open-performance-data [Accessed 5 October 2023].

Performance Review Commission, 2023. *PRC Transformation Support Strategy.* [Online] Available at: https://ansperformance.eu/transformation/

Performance Review Commission, 2023. *PRR 2022 Performance Review Report, An Assessment of Air Traffic Management in Europe,* s.l.: s.n.

Ravenhill, P., Bolic, T. & DG MOVE, 2017. *Interim evaluation of the SESAR Joint Undertaking (2014-2016) operating under Horizon 2020: final report..* s.l.:Publications Office.

Rogers, E. M., 1995. *Diffusion of innovations.* s.l.:Free Press.

Schäfer, M. et al., 2018. *OpenSky report 2018: Assessing the integrity of crowdsourced mode S and ADS-B data.* s.l., s.n., p. 1–9.

Schäfer, M. et al., 2014. *Bringing up OpenSky: A Large-Scale ADS-B Sensor Network for Research.* s.l., s.n., p. 83–94.

SESAR 3 JU, 2023. *DIGITAL SESAR SOLUTIONS CATALOGUE.* [Online] Available at: https://www.sesarju.eu/catalogue [Accessed 29 September 2023].

SESAR 3 Scientific Committee, 2023. *Minutes of meeting, Scientific Committee meeting SC(M)06-2023.* s.l.:s.n.

SESAR Joint Undertaking, 2020. *European ATM master plan:digitalising Europe's aviation infrastructure: executive view: 2020 edition..* s.l.:Publications Office.

SESAR Joint Undertaking, 2023. *SESAR innovation pipeline 2022, Air traffic management research and innovation: 2022 highlights.* s.l.:Publications Office.

SESAR Joint Undertaking, 2024. *SESAR Innovation Days.* [Online] Available at: https://www.sesarju.eu/sesarinnovationdays [Accessed 16 February 2024].

Spinielli, E. & Koelle, R., 2019. *Towards Reproducibility in ANS Performance.* s.l., SESAR Innovation Days.

Strohmeier, M., 2020. *Research usage and social impact of crowdsourced air traffic data.* s.l., s.n., p. 1.

Strohmeier, M. et al., 2015. *Opensky: A swiss army knife for air traffic security research.* s.l., s.n., p. 4A1–1.

Sun, J., Hoekstra, J. M. & Ellerbroek, J., 2020. OpenAP: An open-source aircraft performance model for air transportation studies and simulations. *Aerospace,* Volume 7.

Sun, J., Vû, H., Ellerbroek, J. & Hoekstra, J. M., 2019. pymodes: Decoding mode-s surveillance data for open air transportation research. *IEEE Transactions on Intelligent Transportation Systems,* Volume 21, p. 2777–2786.

Wise Persons Group, 2019. *Report of the Wise Persons Group on the future of the Single European Sky,* s.l.: s.n.

Woodrow, B. I. I. I., 2019. *FAA Chooses Enhanced ADS-C Over Space-Based ADS-B for Oceanic Airspace.* [Online] Available at: https://interactive.aviationtoday.com/avionicsmagazine/august-2019/faa-chooses-enhanced-ads-c-over-space-based-ads-b-for-oceanic-airspace/ [Accessed 31 October 2024].