Open source software and crowd sourced data for operational performance analysis

Reproducibility in operational ANS performance benchmarking

Rainer Koelle
Performance Review Unit
EUROCONTROL
Brussels, Belgium

Abstract—**This paper explores the feasibility of using open data and an open source toolbox for ensuring reproducibility in operational performance analysis of air navigation services. To date the access to operational data in support of research claims or observed performance benefits is limited. Though the majority of related data sets are established and curated by government authorities, open data access is still in its infancy. ICAO promotes a performance based approach. In that light there is stronger interest in performing regional and international benchmarking comparisons. To advance the state of the art it is important to establish common methods and tools. This paper reports on the conceptualization of the benchmarking process as a data analytical workflow and the supporting developments of an open-source data-analytical toolbox. This feasibility study demonstrates the use of the process and open data to provide a framework under which official results can be validated. The benefits of this approach have been demonstrated throughout the on-going regional benchmarking projects. The toolbox developed offers scope to expand the functionality and make operational ANS performance analyses accessible to a wider audience interested in applying harmonized global standards.**

Keywords-Reproducibility; ANS Performance; Benchmarking; Data Analysis

I. INTRODUCTION

Air transportation is a global network and the analysis of operational performance of air navigation services (ANS) within different regions is of interest for policy making and research [1]. Within air navigation the wider access to scientific and analytical data and facts is still limited. To advance the state-of-the-art, results should be made reproducible and more comparable. A possible way to do this is to share tools and data, and allow other researchers and professionals to reproduce the results.

Throughout the recent year the idea of common tools and open data in ATM research and operational validation has gained more visibility. This problem field has been identified in various papers (e.g. [2], [3]). It is recognized that most of the ANS data is collected and curated by government organization. However, the practical implementation of an open data approach is still in its infancy stage. In particular, data sharing policies are limited to specific data products (e.g. public statistics) or purposes (e.g. limited data samples for legitimate research).

One of the key data inputs for operational analysis is the 4D trajectory of a flight. Traditionally, these data are processed for air traffic control purposes and require substantial surveillance data processing systems. Cooperative technologies, like airborne position reporting (e.g. ACARS) and the more recent implementation of Mode-S and ADS-B offer an unique opportunity for affordable signal reception by air traffic control and aviation enthusiasts. The recent years have seen an increase in crowd sourced aviation databases. Modern internet technologies, higher bandwidth, low-cost storage, and community tools support the higher collaboration of aviation enthusiasts. These communities of shared interests range from aircraft photography to scenario development for flight simulation software and flight tracking. Commercial off-springs of the latter are flight tracking providers like FlightRadar24 and FlightAware.

The role and, in particular, the accessibility of the trajectory data was witnessed publically with the loss of Malaysian Airlines flight MH370 in 2014, the downing of Malaysian Airlines flight MH17 over Ukraine in 2014, or the controlled flight into terrain by German Wings flight 4U9525 in the French Alps in 2016. These incidents featured screenshots of the flight paths from FlightRadar24 or FlightAware in the major news channels (e.g. BBC, CNN). As the current license and business model of these providers inhibit the wider use of the crowd sourced trajectory data, open data driven projects have emerged (e.g. OpenSky [4], ADSB Exchange[5]).

This paper addresses the challenges of establishing an open source tool for empirical operational ANS performance analysis making use of crowd sourced data. The contributions of this paper are as follows:

- Conceptualization of the operational benchmarking process as a staged collaborative data analytical workflow;

- demonstration of the feasibility of the chosen approach by using crowd sourced data and an open source tool for the workflow; and

- initial development of associated software implementation for this feasibility study.

This project was performed as a feasibility study to support on-going international benchmarking activities and builds on the feedback received by the involved individuals as well as on the discussion and further development of the GANP KPIs.

II. BACKGROUND

A. State-of-the-art of open data for ANS analysis

Operational performance analysis in ANS research or as part of local or regional performance benchmarking depends heavily on access to associated data. At the time being, the wider access to scientific and analytical data and facts is limited.

For example, research and development programs like the European Single Sky ATM Research (SESAR) or the United States' Next Generation Air Transport System (NextGen) represent consortia. Associated projects validate the proposed operational change or system enabler in form of simulations or demonstrations. The results and benefits are typically hailed in press releases, flyers, and video clips. However, no scenarios and traffic samples are made available publicly to verify the results and benefits. The next phase of SESAR is governed by the Horizon 2020 program of the European Commission. With a view to open data, the Horizon 2020 program puts forward requirements on research projects to establish a data management plan and make the underlying research data publicly available. However, at the same time, grantees have the right to opt-out on the grounds of - inter alia - intellectual property rights or when the objectives of the project might be impaired [2]. Though there is a political realization of the requirement of open data, it is likely that the overall situation will not change.

A similar picture is found when seeking access to operational data. Bourgois and Sfyroeras discusses the role of open data in air transport research [3] by reviewing the underlying data sources of recently published scientific papers in the United States and the European Union. One of the key finding is that about 70% of the data is curated by government organizations. However, existing access and data product policies limit the wider use of the data for operational performance analyses. Access to tailored data sources is notably more facilitated within the United States and driven by the concept of transparent government through the Freedom of Information Act [7]. For example, air traffic related statistics and movement data are available via the Department of Transportations' statistics web-pages [8] or the Aviation System Performance Metrics (ASPM) database of the Federal Aviation Administration (FAA) [9]. In Europe, the EUROCONTROL Agency offers access to comparable data. Nonetheless, the actual access to these data is subject to a screening process to identify legitimate use requests. In a recent change of EUROCONTROL's ATM Data Policy [10], a move towards a more sharing oriented policy was made.

However, practical applications in terms of open data are limited. In 2016, the Performance Review Unit (PRU) launched a project to regularly provide data to support stakeholders interested in validating the published results (http://ansperformance.eu) [11].

B. ANS Performance Benchmarking

ICAO promotes a performance based approach and envisions the future global air navigation system as a seamlessly interoperable system ensuring the efficient flow of air traffic for airspace users during all phases of flight. A key stepping stone in this was the move away from a technology driven perspective on air navigation. The performance based view recognizes that the evolution of the air navigation system is being driven by the expectations of the of the aviation community and supported by technological enablers.

Benchmarking of ANS performance is not a conceptually new topic. Initial work on ANS performance analysis developed in Europe and the US in the late 90s. The collaboration of the FAA and EUROCONTROL in terms of understanding similarities and differences in operational concepts and procedures lead to a series of regional operational comparison reports since 2003 (i.e. most recent comparison report [12]). The work is now governed by a memorandum of cooperation between the United States and the European Union. In late 2016, ICAO endorsed the update of the Global Air Navigation Plan (GANP, Doc 9750) [1]. This version of the GANP comprises a set of 16 key performance indicators (KPIs). The latter are strongly informed by the joint FAA/EUROCONTROL work and the associated practices adopted by a set of air navigation service providers across the world. For example, the civil air navigation services organization (CANSO) published guidance on operational performance indicators in 2013 [13] and 2015 [14], and an initial operational analysis report was published by the University of New South Wales and Airservices Australia in 2012 [15].

Next to the bi-annual US/Europe comparison, the PRU is currently collaborating with several parties in establishing the data collection systems, common interpretation of the KPI algorithms, and the preparation of initial comparison reports. These include:

- Singapore – US – Europe: this tri-partite project focusses on the operational comparison of Changi airport with comparable airports in the US and Europe;

- China – Europe: this project addresses an initial subset of the GANP KPIs for a the top ten airports in terms of IFR movements in China and Europe; and

- Brazil – Europe: the initial benchmarking report focusses on throughput and capacity related GANP KPIs for a subset of airport in Brazil and China.

These projects are characterized by similar challenges. In particular, the establishment of the data collection system for the initial comparison report requires essential effort. Typically data is collected in a rolling forward principle after the identification of the respective data sources. Another constraint

is that for organizational and/or data policy reasons, data cannot be shared openly with other benchmarking parties.

III. CONCEPTUAL BUILDING BLOCKS

A. Open Data for Operational Performance Analysis

The requirement for open data to advance operational performance analyses has been motivated in the previous sections. To avoid constraints due to limitations with the processes surrounding the release and sharing of data amongst the project parties, this feasibility study reviewed currently available data providers.

The data required for operational performance analyses can be broadly categorized in

- Aircraft fleet data – aircraft and operator related data (e.g. registration, type of aircraft)
- Aeronautical data – data for the identification and location of
  - airports and runways
  - relevant airspace related navigation aids, procedure points
- trajectory data – 4D positional information for the construction of flight trajectories

There exists a variety of commercial and non-commercial sources for aircraft fleet and aeronautical data by both, authorities and crowd sourced networks. Non-commercial publications by accredited authorities are, however, typically not in a computer readable format for wider use. For example different states publish their national registered fleet of aircraft or aeronautical information publications in pdf or support the lookup of individual items only. The commercial providers typically provide such databases as part of their premium services, but limit their through prohibitive license agreements.

Crowdsourcing is different from the classical data collection model as the content is collected through the community rather than through traditional suppliers or staff of an accredited organization. Collaboration project for aircraft fleet and (partly) aeronautical data comprise community projects like openflights.org, planespotters.net, and ourairports.com offer data files for aircraft identification and aeronautical base data (e.g. aerodrome and runway related information).

Throughout the recent years several networks of collaborators have emerged for flight tracking purposes. The initial ideas go back to activities based on the reception and processing of HF and ACARS signals. With the introduction of Mode-S and ADS-B there has been a wider application of low cost reception devices and the web-based sharing of the data. A series of youtube videos exist to explain the set-up and operation of respective signal reception dongles. Commercial off –springs comprise FlightRadar24 and FlightAware. These providers use networks of secondary surveillance data receivers and feeders to establish a global air situation. Both providers encourage the further extension of their networks by offering low-cost receivers to interested individuals. But they also benefit from the signals feed by enthusiasts which cooperate with multiple networks. There are two interesting projects, OpenSky [4] and ADSB Exchange [5], that make the data available openly. Both projects differ in the format of the data provision. For the historic data, OpenSky offers a lower level state vector based data repository, while ADSB Exchange provides a json based aggregation of position and meta data based on the Virtual Radar Server AircraftList format. To reduce the overhead in data processing, the data feed from ADSB Exchange was chosen for this feasibility study.

B. Benchmarking - Collaborative Data-Analytical Process

One of the underlying premises of operational benchmarking is the common definition and application of methods and algorithms. The benchmarking process further includes the mutual validation and verification of the results.

This paper addresses operational performance as a data analytical problem. As mentioned above, research, but also benchmarking results, are often presented in dedicated communications (e.g. reports, presentations, flyers, web-pages) without providing deeper insights on how the results have been established while assuring that the presented benefits are correct and tangible or that the specific performance has been observed. This separation of communication and data analysis makes it difficult for other researchers or professionals to verify the findings by reproducing them.

As shown above, one precursor to this research is that data is increasingly ubiquitous - even within the air transportation domain. Accordingly, the focus shifts from the *behind closed doors data world* to the application and validation of common methodologies. Data science in general refers to applying scientific processes to the analysis of data, and extracting knowledge from data. Associated results are considered replicable, if there is sufficient information available to reproduce the results. Integrating the reproducibility into the data analytical workflow can be understood as a data-analytical program [16] as presented in Fig. 1.
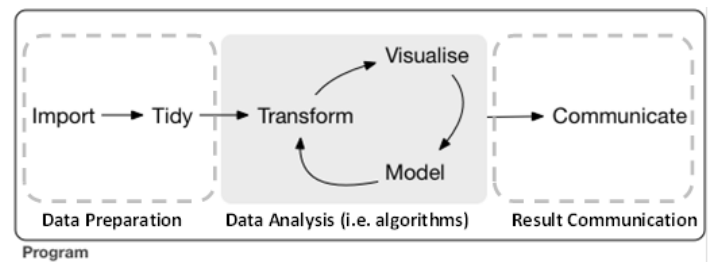

Figure 1. Data-Analytical Process

The workflow of Fig. 1 can be broadly broken down into three stages:

- data preparation – the actual identification and import of the data collected for the respective study, and its transformation to support the further analysis (i.e. data cleaning, tidy data);
- data analysis – the application of the agreed common algorithms for performance benchmarking, including the harmonized application of parameters; and

- result communication – the generation of reports, presentations, web-pages, etc. to reach out and inform about the results of the analysis.

For the domain of operational ANS performance, the principle of reproducible data analysis can be formulated as two mutual dimensions:

1. making the methodology and empirical work (i.e. data and analysis code) used for establishing the findings available; and

2. ensuring sufficient transparency in the process for an independent researcher or professional to reproduce the findings.

Arguably there are limits to the access to underlying source data. For example, the data volume might prohibit further dissemination or its collection falls under intellectual property rights or restrictive license agreements. For that purpose the creation of an analytical data set as the output of the data collection, gathering, and cleaning process stage can be understood as sufficient in the sense of reproducible research. The article Spies in the Skies by Peter Aldhous and Charles Seife [17] is a good example for a reproducible analysis of the US government's airborne surveillance program. Though the license restrictions did not allow for the publication of the underlying aircraft position data, the authors made available the analytical data set and documented their analysis.
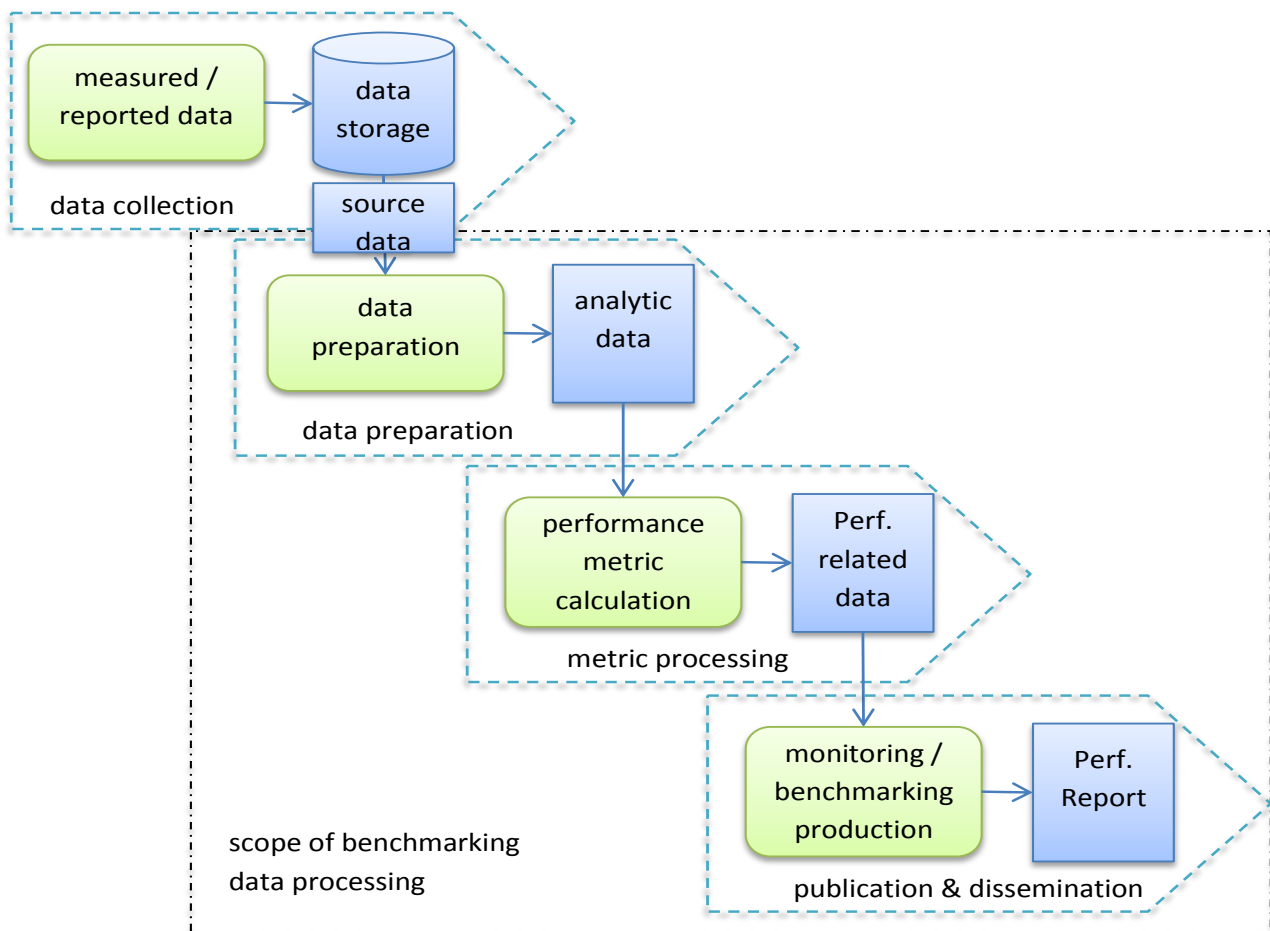

Figure 2.   Data-Analytical Process

IV. RESULTS

The contributions and results of this project can be broadly broken down into the following categories:

- Collaborative benchmarking process;

- Application of common methods and algorithms through the development and use of an open source data-analytical toolbox;

- Utility of the open data in terms of coverage; and

- Summary of the numeric results of the feasibility.

A. Collaborative Benchmarking Process

Given the nature of the benchmarking process and associated requirements to ensure transparency and reproducibility, the general data analytical process presented in Fig.1 was adapted as presented in Fig. 2. In particular a set of clearly defined stages were implemented with associated process data sets:

- Source data – source data forms the input data to the analysis. The term source is not to be understood as raw data, but can refer to the data set that is extracted from the data collection system for the purpose of the benchmarking study. It follows that in the absence of data provided by a benchmarking partner, open data can serve as a valuable input to the benchmarking process.

- Analytical data – (a set of) data set(s) that form the input to the analytical process. In particular, these data are free from data artefacts and can be processed in accordance with the chosen algorithm

- Performance related data – the output oriented data set(s) that document the result of the processed analytical data.

The process depicted in Fig. 2 and its inherent data abstractions provide a framework for collaborative benchmarking by abstracting data and analysis (i.e. common algorithms) and enabling reproducibility. The latter is achieved as prohibitive data samples (e.g. volume, use licenses) can be made available as a representative sample. For example, in a study for several airports, the data from two airports, or a monthly sample from the majority of airports is sufficient to validate the performance related data. Using further open data augments the provided analytical data set and allows for the determination of estimates.

This process was applied in the regional benchmarking processes. In particular, PRU shared representative samples of the analytic data with benchmarking partners. Based on the open source implementation of the data analysis, stakeholders were able to validate the results (performance related data) shared by PRU. The continued application of these principles and abstractions was then adopted by the benchmarking partners.

B. Common Methods - Open Source Data Analytical Toolbox

One of the design criteria of this feasibility study was to investigate the use of free open source software for the implementation of the data analytical transformation and presentation of the data. There is a variety of open source software available. Furthermore, different commercial product providers offer free versions of their products as appetizers (e.g. Tableau). The trade-off for the latter class was that typically the capabilities of the open version was limited in terms of data volume, recognized data formats, and output limitations.

Amongst the reviewed open source products R was chosen [18]. R is an open-source software environment for statistical analysis and graphics. One of the key features of R is its extensibility. Researchers and developers make their tools, i.e. R software implementations, available as so-called packages. There are many packages available today and are hosted either by CRAN or more specific scientific discipline sites, e.g. bioconductor. On-going package development and maintenance work is often hosted on github [19]. This offers a rich universe for data analysis in general and an opportunity for ANS performance related work.

R is supported by another open source IDE, RStudio [20]. RStudio integrates the R engine into a toolset for the data analytical process providing functions for code production, interactive R session handling, and supporting features like graphics and output viewers. It also supports the interface with commonly used software version control and archiving solutions (e.g. github, bitbucket).

The work presented in this paper makes use of the following packages:

- rmarkdown [21] - a R-dedicated mark-up language that support the combination of text, R code chunks, and graphics. rmarkdown documents can be parsed and rendered into formatted output, e.g. html, pdf, or MS Word output. This eases the result communication tremendously. Another benefit of rmarkdown is that it allows to fully reproduce the output. This means, that once the respective code chunks have been defined, the injection of the same data set will reproduce the identical results. Moreover, the scripts ensure the application of the same common method to other data sets (e.g. an open data set).

- tidyverse - one of the most influential developers of packages for R/RStudio is Hadley Wickham [16]. tidyverse is a collection of packages for data analysis. This ranges from tools for the data loading to data transformation, e.g. dplyr, and visualization, e.g. ggplot2. One of the benefits of the tidyverse packages is that these follow a verb-based programming logic. This reduces the learning curve.

Another benefit in choosing the R/RStudio/rmarkdown ecosystem is the tight integration of reproducibility and output generation. At the time being rmarkdown supports a variety of output formats, such as pdf, MS Word, and html5 based web-pages or slides.
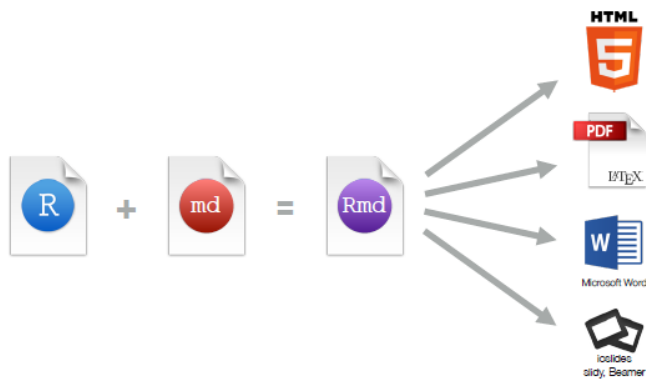
Figure 3.   rmarkdown – output generation

This feature was heavily valued as part of this feasibility study. Changes to the code base could be reflected immediately in the generated output (following the required re-parsing of the rmarkdown document). The variety of output formats offers a great benefit, as code chunks can be reused to produce outputs like publications or slides. For example, a draft version of this paper was produced with making use of the performance related data produced through the projects in combination with the accompanying text.

One of the key benefits of using R is its package system. Following up on the approach chosen in this paper, the different implementations of data preparatory activities and data analytical transformations of the data are made available as a package. The package is hosted on github to offer other researchers and professionals the opportunity to contribute and refine the respective metric implementations. This includes a complete example processing chain from the download of the respective open data to the implemented data cleaning and KPI metric calculations.

C.   Utility of Open Data – Benchmarking the Benchmark

One key point of this feasibility study was the processing of open data to support the international benchmarking activities. The use of open data allowed for a targeted discussion of data requirements and the numerical analysis. In particular, benchmarking partners facing restrictive data sharing policies or in the process of establishing their data collection capabilities, appreciated the '*learning while analyzing*' approach. This proved extremely beneficial.as efforts were invested in understanding and discussing initial results pointing out differences in air traffic control and management procedures. Based on the open source analysis tool, it was also possible to jointly develop and implement supporting metrics to inform the specifics of the benchmarking project.

The collaborative process in combination with the open data allowed further for a qualitative benchmarking of the used data sources. For example, Fig 3 shows the trajectory data coverage for Europe based on the data feed of ADSB Exchange. Geographically, this data source covers most of Europe. The major traffic flows in Europe are visible. Noteworthy are areas of no coverage, e.g. the southern part of
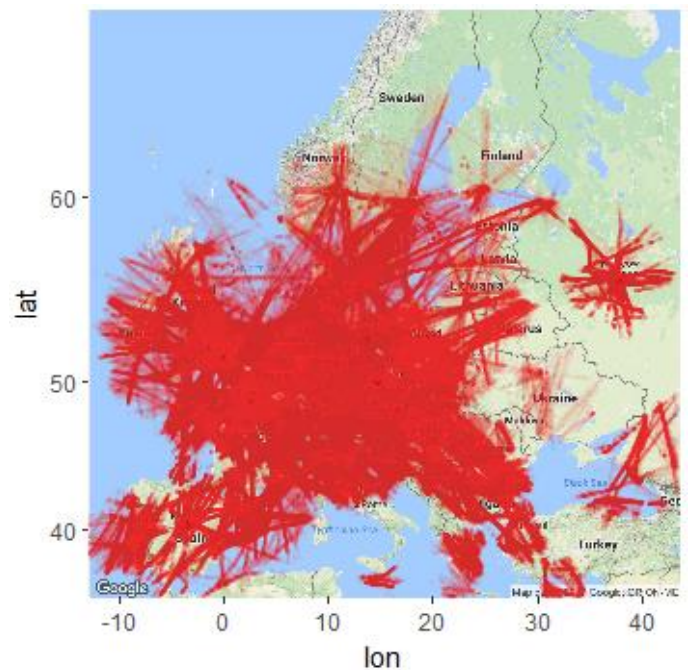

Figure 4.   ADSB Exchange Data Coverage in Europe

Italy or the intermediate segment for traffic to/from Moscow and the eastern stretch of Europe.

The trajectory data does not provide ground coverage at all European airports. This equally applies for other regions of this planet. However, in most of the cases, the trajectory data provides a sufficient coverage for observing aircraft on final or touching down. This limits the application of the data to throughput and arrival/departure phase indicators. Nonetheless, the estimates for the departure and landing times determined on the basis of data from the initial climb out and final approach showed a good match with the operational comparison data.

All trajectory data providers point towards enhancing their coverage by inviting feeders to share their data with their networks. It is therefore anticipated that the overall coverage will improve over time and, hence, increasing the utility of these sources for regional performance analyses.

Interestingly, the aircraft fleet and aeronautical data used for this feasibility study and on the basis of the chosen benchmarking airports proved of high accuracy. As part of the joint analysis, all partners validated the open data used for the study. This shows that the general premises of collective quality assurance of crowd sourced data is valid.

D.   Feasibility Study

This feasibility study focused on the use of open data and an open source software toolbox for operational ANS performance analysis. The study revolved around the scope established in the on-going international benchmarking projects.

For that purpose, the source data sets have been accessed and downloaded from the given sources. The data for the production of flight trajectories represents the bulk of the

study. The R environment offers a variety of packages and functions for the download and reading of data files. Given the size of the ADSB data, i.e. on average 2.2GByte for a day, the download from ADSB Exchange took about 45-60 minutes on a regular office internet connection per study day. An initial set of one month of data was downloaded (i.e. July 2016). Other data was attainable as a simple download of the respective data files.

As mentioned above, the ADSB Exchange data is stored in daily files represented in the AircraftList format. This required a further pre-processing of the downloaded data files to extract the actual 4D position information. This pre-processing step allowed additionally to build a fleet data base and an airport look-up table. Both side-products were useful for the validation of the data completeness as missing values in terms of aircraft identifiers (i.e. unique hex ID, aircraft registration, type) could be imputed.

The following table III summarizes the level of coverage and completeness of the 4D trajectories established through the ADSB data feed. In general the data feed is fused from different feeders and as such the data per aircraft is augmented throughout the processing. Nonetheless, there seems to be no dedicated quality assurance process when different feeds are combined, i.e. it is possible that the same aircraft is tracked by multiple sources with varying reference data sets.

TABLE I.   SUMMARY TABLE OF FEASIBILITY STUDY

Study airport	trajectory based results for 1 month of data (July 2016)		
	complete-ness	*coverage*	*add.ASMA time (40NM)*
EGLL	98%	84%	7.2 min/arr
EHAM	99%	89%	0.90 min/arr
LFPG	98%	87%	0.29 min/arr
EDDF	97%	85%	1.32 min/arr

As part of this feasibility study an initial additional time in the arrival and metering area (ASMA) was established. For this purpose the algorithm determines a 40NM crossing point from the airport with the trajectory. The actual landing time is estimated based on the data points measured on the final approach / before touchdown. An appropriate reference time for the approach segment is determined as the 20th percentile over the travel times observed. The results are presented in Tab. III and range under 10% difference from the data published by the PRU for the respective airports.

Considering the implementation of the algorithm in accordance with the ICAO GANP KPI (which differs from the PRU algorithm) and the coverage level of above 85% on average, the metrics calculated on the basis of the open data can be considered adequately close to the official results. The study also revealed that 85% coverage appears to be a reasonable lower bound of coverage that supports the calculation of the metric. As a side result, the coverage provides also a quantification of the current level of implementation of the ADS-B related fleet equipage requirements. The deadline for ADS-B installation for aircraft operating in the US and Europe is 2020. As the chosen set of airports represent major hubs it can be inferred that the majority of international operators is already compliant with the requirement.

V. CONCLUSIONS

This paper revolves around establishing an open-source analysis capability to support operational ANS performance benchmarking. In particular, the feasibility study addressed the integration of crowd sourced open data to support the data analytical task in the absence of sharable data amongst benchmarking partners. The feasibility study was driven by ensuring reproducibility in terms of developing common methods and applying common tools. For that purpose the benchmarking process has been abstracted as a data analytical process with defined stages of data sets. The latter allows for the definition of tailored data sharing mechanism amongst benchmarking partners to accommodate potentially restrictive data sharing policies. It has been proven that the process and its implementation served as a communication tool and ensured the timely provision of a representative data sample on which the results of all parties could be validated.

The scope of the project is currently linked to on-going activities of the EUROCONTROL PRU and comprises work under the US/Europe comparison study, and initial benchmarking activities with Singapore, Brazil and China. These projects tackle subsets of the ICAO GANP indicators and commonly agreed further metrics. The feasibility study aimed for a development of a wider set of these metrics to provide a wider toolset for interested parties as part of the R package system.

The approach and data sources presented show a sufficient level of coverage. However, ground coverage is not available for all airports on a global level. Accordingly, these metrics – though the respective indicator metric implementation is made available – cannot be established without augmented data samples from accredited authorities. This limitation will eventually evaporate as the coverage of the feeder networks will improve.

This paper focused on the initial work of utilizing R/RStudio for operational ANS performance benchmarking. The results obtained serve as a blueprint for the further development of a supporting benchmarking package for R/RStudio. Most recent developments for rmarkdown support the integration of code into interactive web-pages and template engines. The latter supports the production of performance dashboards with minimal additional development effort. An initial proposal for such a dashboard is currently under development. It is also foreseen to maintain and enhance the analytical functionality with future benchmarking projects. This development will be shared publicly via github to ensure proper outreach and active participation by the performance community or interested researchers. Initial coordination work is on-going to explore the possibility of establishing a worldwide reference trajectory for performance purposes. The benefit of this approach and a potential collaboration under the umbrella of ICAO is under discussion.

REFERENCES.

[1] International Civil Aviatio Organization (ICAO), Doc 9750, Capacity & Efficiency, Global Air Navigation Plan 2016-2030, Fifth Edition, 2016.

[2] J.M. Hoekstra and J. Ellerbroek , "BlueSky ATC Simulator Project: An Open Data and Open Source Approach", 7th International Conference on Research in Air Transportation: Philadelphia, USA, 2016, uuid:d1131a90-f0ea-4489-a217-ad29987689a1;
BlueSky simulator: https://github.com/ProfHoekstra/bluesky

[3] Marc Bourgois and Michael Sfyroeras, Open data for air transport research: dream or reality?, 2014,
URL: http://www.opensym.org/os2014/proceedings-files/p303.pdf.

[4] Matthias Schäfer, Martin Strohmeier, Vincent Lenders, Ivan Martinovic and Matthias Wilhelm. "Bringing Up OpenSky: A Large-scale ADS-B Sensor Network for Research", Proceedings of the 13th IEEE/ACM International Symposium on Information Processing in Sensor Networks (IPSN), pages 83-94, April 2014,
The OpenSky Network: URL: https://opensky-network.org.

[5] ADSBexchange, URL: http://www.ADSBexchange.com.

[6] European Commission, Open Research Data in Horizon 2020, infographic, July 2016.

[7] United States Government, The Freedom of Information Act 5 U.S.C § 552, as amended by public law, No 104-231, 110 Stat. 3048, 1996.

[8] United States Department of Transportation, Bureau of Transport Statistics, Airline Data TranStats,
URL:https://www.rita.dot.gov/bts/data_and_statistics/transtats/index.html.

[9] Federal Aviation Administration, FAA Operations & Performance Data, URL: https://aspm.faa.gov/.

[10] EUROCONTROL, Agency ATM Data Policy and Regulation on ATM Data Protection and Public Access. Note de Service / Office Note No 18/12, May, 2012.

[11] EUROCONTROL, Performance Review Unit (PRU), PRU Portal, URL: http://ansperformance.eu.

[12] EUROCONTROL on behalf of the European Union and Federal Aviation Administration Air Traffic Organization, 2015 Comparison of Air Traffic Management-Related Operational Performance: U.S./Europe, August 2016.

[13] Civil Air Navigation Services Organisation (CANSO), Methodologies for Calculating Delays/Improvement Opportunity Pools by Phase of Flight, May 2013.

[14] Civil Air Navigation Services Organisation (CANSO), Recommended Key Performance Indicators for Measuring ANSP Operational Performance, March 2015.

[15] Tim Rees, University of New South Wales and Airservices Australia, Analysis of Australian ATM-Related Operational Performance, Version 1, October 2012.

[16] H. Wickham and G. Grolemund, R for Data Science, Import, Tidy, Transform, Visualize, and Model Data, O'Reilly Media, December 2016.

[17] Peter Aldhous and Charles Seife, BuzzFeedNews, Spies in the Skies, April 2016, URL: https://www.buzzfeed.com/peteraldhous/spies-in-the-skies?utm_term=.qjBppEG1D0#.wa8kk4qryp.
URL: github-repository: https://buzzfeednews.github.io/2016-04-federal-surveillance-planes/analysis.html.

[18] R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

[19] Github, community tools for software development, URL: https://github.com/.

[20] RStudio Team (2016). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA URL http://www.rstudio.com/.

[21] JJ Allaire, Joe Cheng, Yihui Xie, Jonathan McPherson, Winston Chang, Jeff Allen, Hadley Wickham, Aron Atkins and Rob Hyndman (2016). rmarkdown: Dynamic Documents for R. R package version 1.3. https://CRAN.R-project.org/package=rmarkdown.

AUTHOR BIOGRAPHY

Rainer Koelle heads the Safety and Quality of Service section of the Performance Review Unit at EUROCONTROL, Brussels. He holds a PhD from the University of Lancaster, 2013, in Communication Systems. His research interests include data analytics for operational performance analysis.