This project has received funding from the European Union's Horizon Europe Programme

RESEARCH PROJECT EASA.2022.HVP.01

D-3.3.5 – DEVELOPMENT OF INDUSTRY-AGREED FDM ALGORITHMS AND LOGICS

Digital transformation - Case studies for aviation safety standards – Data Science Applications (DATAPP)

An Agency of the European Union

Disclaimer

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Union Aviation Safety Agency (EASA). Neither the European Union nor EASA can be held responsible for them.

DELIVERABLE NUMBER AND TITLE:	DATAPP D-3.3.5 Development of industry-agreed FDM algorithms and logics
CONTRACT NUMBER:	EASA.2022.HVP.01
CONTRACTOR / AUTHORS:	ALG/C. Cuesta, A. Feliubadaló, A. Olivera, A.Cabeza & A. Bujor
IPR OWNER:	European Union Aviation Safety Agency (EASA)
DISTRIBUTION:	Public

APPROVED BY:	**AUTHOR**	**REVIEWER**	**MANAGING DEPARTMENT**
	ALG	**Núria Alsina**	**ALG**

DATE: 16 July 2024

SUMMARY

Problem area

Data stays at foundation of decision-making, accelerating the digital transformation across industry. Strong data systems and new technology have been embraced in aviation with significant changes to the traditional working processes, business models, standards and regulations. In this context, EASA faces new challenges on what the required changes in safety standards and regulations are needed in response to the introduction of innovative solutions and processes. Anticipating what is to come in the industry in the field of data science applications is key to make sure safety levels are maintained without slowing innovation down.

The objective of this project is to identify and assess relevant changes to the existing aviation safety standards to support the deployment of the digital solutions under three case studies:

- Case Study 3: Flight training data for EBT/CBTA (Evidence-Based Training / Competence-Based Training and Assessment).
- Case Study 4: Digital fuel management.
- Case Study 5: Flight data models for safety.

The project aims to provide a comprehensive evaluation of benefits, constraints, standardisation and deployment issues, including the recommendations for adjusting safety regulations and related standards, and how new digital technologies could contribute to addressing the identified issues.

Description of work

This report belongs to the task “T-3.3 Training material” of the “Digital Transformation – Case Studies for Aviation Safety Standards” project (EASA.2022.HVP.01- Horizon Europe Project). The purpose of this deliverable is to provide dissemination material designed to concretize some of the solutions identified within the project's context, particularly those that could represent potential quick wins. This material aims to offer initial and independent reflections from the consultant, serving as a foundational resource for future initiatives by the Agency and the industry. By outlining these actionable insights, the document seeks to stimulate further exploration and implementation of effective strategies, ultimately contributing to the advancement and innovation within the sector. Specifically, this document provides recommendations for the standardisation of the definition of FDM events, which would potentially contribute to the improvement of the operators programmes' efficiency, to facilitate industry-wide blind benchmarking, and to enhance and foster best-practice sharing across the industry.

Results and Application

The report delves into one of the solutions proposed in the context of the project, providing further details and a series of recommendations that could be applied to achieve an effective implementation of the solution. All of this is collected and provided in the form of training materials. Such training materials are intended to be used at EASA's discretion, for instance by including it in dissemination documents or in guidance material to help in the potential implementation of the solutions by the stakeholders. Thus, the output of this material

provides additional information to EASA to support their decision on the evolution of the solutions proposed in the context of this project.

CONTENTS

ABBREVIATIONS

ACRONYM	DESCRIPTION
ACARS	Aircraft Communication Addressing and Reporting System
AMC	Acceptable Means of Compliance
AOC	Air Operator Certificate
ATO	Approved Training Organisations
AU	Airspace Users
CAT	Commercial Air Transport
CBTA	Competency-Based Training and Assessment
CS	Case Study
D4S	Data4Safety
EAFDM	European Authorities coordination group on Flight Data Monitoring
EASA	European Union Aviation Safety Agency
EBT	Evidence-Based Training
EOFDM	European Operators Flight Data Monitoring forum
FDAU	Flight Data Acquisition Unit
FDM	Flight Data Monitoring
FSF	Flight Safety Foundation
GM	Guidance Material
IATA	International Air Transport Association
KPI	Key Performance Indicator
MRO	Maintenance, Repair and Overhaul
NAA	National Aviation Authority
SMS	Safety Management System
SPI	Safety Performance Indicators
SRM	Safety Risk Assessment
UC	Use Case

1. Introduction

1.1 Scope of the document

This report represents one of the deliverables under the task “T-3.3 Training material” of “Digital Transformation – Case Studies for Aviation Safety Standards” project (EASA.2022.HVP.01- Horizon Europe Project). D-3.3 is complemented by 5 (five) individual deliverables covering the different case studies of the project, presented in Table 1-1 below.

Table 1-1 List of deliverables as D-3.3

Deliverable		Title	Case Study
D – 3.3	D-3.3.1	Standardised metrics and methods for instructor concordance assurance	CS3 - EBT
	D-3.3.2	Untapped benefit of fuel reduction schemes: Reviewing the NPA-2016-06(A) economic impact assessment	CS4 - Fuel
	D-3.3.3	Recommendations on assurance framework for analytical development and approval of fuel schemes	
	D-3.3.4	Training requirements for FDM analyst	**CS5 - FDM**
	D-3.3.5	**Development of industry-agreed FDM algorithms and logics**	

Within the context of the DATAPP project, the D-3.3.x deliverables are intended to provide dissemination material designed to concretise some of the solutions proposed during the project, particularly those that could represent potential quick-wins. Such materials are intended to be used at EASA's discretion, for instance by including it in dissemination documents or in guidance material.

This deliverable “D-3.3.5 Development of industry-agreed FDM algorithms and logics” focuses in providing recommendations for the standardisation of the definition of FDM events in a context of collaboration and agreement between the main industry stakeholders through mechanisms such as national collaborative programmes. Through such standardised event definitions, any operator that has to implement an FDM programme would have further guidelines in this regard. Besides contributing to an improvement in the efficiency and effectiveness of individual operators' programmes, this could also greatly facilitate effective blind benchmarking exercises aimed at drawing conclusions on an industry-wide basis as well as enhance best-practices sharing between the main stakeholders.

The present document is structured as follows:

- Section 1 as an introduction presenting the background scope of the document.
- Section 2 presents an overview of the proposed solution, including discussion on potential approaches to favour development of industry-agreed standardised FDM event definitions.
- Section 3 includes the conclusions where the results of this deliverable have been summarised.
- Section 4 lists the reference material that have been used as a reference for developing this document.
- Annex A provides an illustrative example of FDM event definition card.

- Annex B provides high-level considerations for selection of minimum representative datasets in the context of collaborative analytical works of FDM data.

2. Development of industry-agreed standardised FDM events definitions

2.1 Background

2.1.1 Why we need to act – Issue / rationale

In the context of the research conducted under the DATAPP project, multiple organisational, technological and procedural constraints affecting the effectiveness of FDM programmes were identified, with a wide disparity in programme maturity among European operators. These limitations were aggravated in those operators with smaller fleets, with difficulties in investing in technology and human resources. The flexibility provided at the regulatory level for the implementation of FDM programmes, focused on the promotion of such programmes through a framework adaptable to the reality and needs of each operator, also leads to the disparity in their implementation and with the effectiveness of the programmes. Such limitations are aligned with the rationale of the amendments proposed by EASA in NPA 2024-02, based on feedback from standardisation inspections (in EASA's non-public Annual Standardisation Reports for 2019 and 2020) and evaluation of the European Operators Flight Data Monitoring forum (EOFDM) best practice documents.

AMC1 ORO.AOC.130 on flight data monitoring does not include performance objectives to ensure minimum effectiveness of the FDM programmes. As indicated in the NPA 2024-02: “in several visited operators, FDM was not adequately used by the operator, which hampered its identification of operational hazards and therefore its safety risk management process”. Among the minimum performance conditions, one of the major limitations to the effective evolution of FDM programmes lies in the difficulties in capitalising and maintaining analytical knowledge, namely in-flight data collected and FDM algorithms implemented in the software used. To maintain this knowledge over time, the operator should document information on the data source (which aircraft sensor or system) and the performance of flight parameters (recording resolution and recording rate). This information should be documented as it is essential for ensuring that flight parameters used by the FDM algorithms (FDM event algorithms and FDM measurement algorithms) that are programmed in the FDM software are adequate. This documentation should be controlled and include the history of modifications, to retain knowledge of changes made despite the evolution of the fleet and FDM staff changes. The same would apply for information on the FDM algorithms, ensuring that these algorithms are adequate for the aircraft model, type of operation, SOPs, etc., or, if necessary, to perform (or request) an adaptation of these FDM algorithms.

The safety risks that operators face during a flight can be either common across the industry, or specific to their operation and operational context. Still, the definition of events and flight phases, the algorithms that represent these definitions, and even the parameters used, are not standardised, with each operator and software vendor using their own. It is worth mentioning that the European Operators Flight Data Monitoring forum (EOFDM) Working Group B has developed relevant guidance material[1] for the implementation of flight data monitoring precursors, providing details on methods to monitor key precursors of incidents. However, to date the working group and its outputs are not widely disseminated throughout the industry. Amendments proposed to the Air Ops regulations within the NPA 2024-02 explicitly provides reference to such document, as it could be in the GM2 ORO.AOC.130, what could promote its diffusion and use for FDM programmes in the EU.

---

[1] EUROPEAN OPERATORS FLIGHT DATA MONITORING FORUM (EOFDM) WORKING GROUP B, “Guidance for the implementation of FDM Precursors”

As a result, today, many operators still use predefined algorithms that are provided with the FDM software, and they unfortunately perform no or limited adaptation of these algorithms. This results in, among others, four main issues:

Difficulties in adapting FDM events definitions or creating new ones

Operators and software vendors often struggle to adapt existing FDM events, flight phases and thresholds or define new ones due to the lack of guidance available. The limited availability of information on industry's best practices, specific definitions and algorithms hampers the development and customisation of FDM programmes tailored to specific aircraft and operational contexts. Although there is reference material created by collaborative working groups, these do not usually detail specific pseudo-codes for measuring conditions, provide guidance on basic pre-processing rules to ensure the effectiveness of events and/or conditions for defining thresholds for all events. On top of that, some aeroplane operators may not necessarily consider the safety issues that appear to be the most frequent and/or severe when considering national or European accident statistics, as a priority. However, no operator is immune to these safety issues; if they are not properly monitored it is just a question of time that the operator experiences an event related to the top safety issues.

Thus, establishing clear and comprehensive event definitions would provide the necessary framework and guidance for creating and adapting FDM events, enhancing the overall quality and effectiveness of FDM programmes.

Implementation of FDM programme by operators of smaller aircraft

Regional aircraft with a MCTOM of less than 27 000 kg, and particularly those equipped with turboprop engines, do not currently fall under the provision to be covered by an FDM programme. Many operators, however, choose to include these aircraft into their existing FDM programmes to benefit from both safety intelligence and the kick-start to their flight data-based programmes more broadly.

In many cases, the investment dedicated to FDM programmes is nevertheless lower, and the tools available for their operation are less advanced and/or the effort invested in adapting the FDM event definitions is less. As a result, such users are currently limited by the lower quality of events, measurements, flight phases identification and other flight data-related algorithms, when compared to those of aeroplanes with MCTOM > 27 000 kg, resulting from a lower understanding of the specificities of these aircraft and their operation. Given the potential extension of scope for aeroplanes requiring an FDM programme (ICAO plans to extend it to those with an MCTOM of more than 15 000 kg and a MOPSC of more than 19 passengers), this limitation may gain in relevance in the future. In this context, a standardisation framework for FDM events would help to make the effective implementation of FDM programmes for these operators more progressive and smooth.

Reduced effectiveness of best-practices sharing

Some States and organisations have set up so-called large data exchange programmes, under which they gather very large amounts of data (including FDM data) provided by many operators and by other industry stakeholders, which are then centrally processed and analysed. Participation in a large data exchange programme may bring various benefits for an operator, such as being able to compare its safety performance with that of comparable operators or getting access to other types of data (weather, traffic, etc.) or to advanced data integration capabilities. In addition, in the case of an operator with a small fleet producing small amounts of flight data that do not allow for reliable identification of trends, joining a large data exchange programme might help to overcome this limitation.

However, the lack of a standardised framework of events definition implies a major limitation for these programmes. More specifically, the variability in event definitions reduces the applicability of best practices shared among operators since the results of events and measurements cannot be directly compared and/or adopted within the operator FDM programme. Also, when operators exchange information and best practices on FDM event definitions, the effectiveness of these communications is often diminished due to the existing differences in definitions and implementation. This issue can also impact the willingness of operators to actively participate in industry forums and exchanges. In that regard, standardised event definitions would ensure that best practices are relevant and applicable across different FDM programmes, fostering greater collaboration and knowledge sharing within the industry.

Challenge on performing benchmarking exercises

The analytical ability to compare different dimensions of safety performance against other industry users is one of the major benefits of collaborative programmes. In particular, this often translates into 'blind-benchmarking' capabilities, where through analytical rules of data anonymisation, companies can compare their performance while protecting the potential identification of any other company. This is a common core product delivered by data-exchange initiatives (e.g. EASA Data4Safety, FAA ASIAS). In the context of these initiatives, which involve the exchange of data at the highest granularity and pooling of analytical effort, the development of these capabilities is already a challenge in itself, requiring an effort in the development of harmonised and agreed standards for the generation of metrics to feed the benchmarking capability. In turn, as will be discussed later, this work produces harmonised best-practices that could serve for the promotion of standardised FDM metrics and practices at industry level. The challenge is not only technological, but affects data governance as well: how is data shared? Who can analyse it? At what granularity and with what temporal dimension? Who can see what and under what conditions of anonymisation?

In those programmes with pooling of FDM event results pre-processed by each individual operator, the benchmarking exercise is even more difficult, given that these cooperations are commonly governed by more strict data governance and confidentiality policies, which directly affect the consistency of the comparison exercises. Due to varying event definitions, in this context direct benchmarking of event trends and rates becomes challenging. Without standardised definitions, comparing safety performance across operators becomes less meaningful and representative, reducing the utility of benchmarking exercises. This is why in these contexts a significant effort is usually devoted to the redefinition of standardised FDM events, based on a harmonisation of triggering conditions, flight parameters pre-processing and thresholds, considering the operational granularities necessary for the effectiveness of the event (airport, aircraft type...) but without considering the particularities of the operator.

In that regard, standardised event definitions enable relevant comparisons on the same basis, allowing operators to identify safety trends, share best practices, and drive continuous improvement in safety performance.

2.1.2 What we want to achieve – Objectives

This proposed solution aims at contributing to achieving the overall objectives of improving FDM programme effectiveness at operator and large exchange programme levels, by addressing the issue on the lack of standardised FDM events definitions & documentation framework. Standardising the definition of FDM events is challenging due to the historical precedent of each airline independently defining these events. To address this, the proposed solution involves leveraging on the main stakeholders' collaboration across the industry to achieve a soft-level standardisation of FDM event definitions, similar to ICAO's Global Air Navigation Plan (GANP) where multiple variants are acceptable for the same Key Performance Indicator (KPI). This approach allows for flexibility and adaptability. A reference in that regard could be the Data4Safety Programme, which could potentially offer a comprehensive solution to the challenges posed by the lack of standardisation in FDM programmes.

The objective of standardised FDM events is NOT to completely standardise the FDM programmes of aircraft operators. This is not desirable, as:

- The regulator cannot define an exhaustive list of FDM events, given the variety of aircraft models, types of operation and airfields. The definition of an appropriate FDM event set will vary according to the nature of the operation.

- The aircraft operator is responsible for managing its own SMS, including its safety data sources, such as the FDM programme; it is a general safety management principle that the data collection should be adapted to monitor particular risks identified. FDM programmes need to account for their own operational safety priorities.

However, standardised FDM events are foreseen to represent a subset of the FDM events that are relevant for the prevention of the categories of occurrence identified as priority by the EPAS. These standardised FDM events need to be programmed according to a common definition, as far as practicable.

The specific objectives of this proposed solution are to:

- Identify possible working approaches to get such standard definitions used at the EU level and reflection on potential means to achieve widespread use of standard FDM events definitions
- propose a framework for the standardised definition of FDM events, complementing the recommendations made by the EOFDM WGB in the guidance material produced
- Describe key analytical methods for implementing standardised FDM events in collaborative data sharing frameworks

2.2 How we could achieve it – Overview of the proposed solution

The overall purpose of this solution is to build a framework for the development of standardised FDM events definitions, agreed at industry level, materialised as guidance material that can be adopted by all kinds of operations and fleet. On top of that, such standardised event definitions could also potentially ease the collection and comparison of data from various operators, in the context of FDM collaborative analysis initiatives and/or large data-exchange collaborative programmes. As captured in the proposed amendments to the NPA 2024-02, documentation on the algorithms to produce FDM should be sufficiently detailed to capture the trigger logic, event trigger thresholds and the required parameters.

The standardisation of the definition of FDM events is not a novel work in the industry, and there are already several relevant works within the framework of collaborative programmes for such harmonisation. The European Operators Flight Data Monitoring forum (EOFDM) Working Group B has developed relevant guidance material for the implementation of flight data monitoring precursors, providing details on methods to monitor key precursors of incidents. Amendments proposed to the Air Ops regulations within the NPA 2024-02 reference such document, as it could be in the GM2 ORO.AOC.130. Similar work is being captured in the 'Flight Data Metric definition' elaborated by the Commercial Aviation Safety Team (CAST)/ICAO Common Taxonomy Team (CICTT). The European Authorities coordination group on FDM (EAFDM) published a document proposing a set of standardised FDM-based indicators, being a subset of the FDM events, which are meant to capture potentially unsafe situations during aircraft operation. The document contains proposed definition cards with the relevant fields to be considered and which are further explained in the “Reference event definition card” section. This direction should be further pursued in the context of the event definitions, which should be defined in a more detailed and standardised way.

Building on the aforementioned works, the objective to be pursued would be to complement the work carried out under the recommendations presented under the EOFDM working group for the definition of precursors, incorporating the standard structuring proposed by the EAFDM working group and deepening in particularities that may favour comparable analytical results.

In this regard, an additional relevant reference to be considered is the guidance material prepared for the implementation of an FDM event standard for the identification of Unstable Approaches, produced under the framework of the EASA Data4Safety programme. Such process is covered in the “Guidance for identifying unstable approach with flight data” document. This guidance document provides comprehensive indications for air operators to detect and analyse unstable approaches using flight data. The document begins by defining an unstable approach, also specifying criteria based on flight parameters such as airspeed, altitude, rate of descent, and configuration. Moreover, it establishes harmonised thresholds for these parameters to ensure that deviations indicating instability are identified in a standard way at programme level. By presenting the assumptions, the applied methodology, the considerations and the lessons learnt derived from the development of an unstable approach detection algorithm, the document helps to ensure that unstable approaches are accurately detected and analysed under harmonised criteria.

Considering the work done and the existing collaboration frameworks, two possible approaches, potentially complementary, are therefore identified for the development of industry-agreed FDM events standard definitions:

2.2.1 Approach #1 – Dedicated working group further developing FDM events definitions

As mentioned above, there are multiple collaborative groups focused on the development of similar materials, consisting of different types of stakeholders, as well as operators, manufacturers and FDM software providers. In this sense, EASA could rely on the groups that currently support FDM as the EOFDM, which facilitates collaboration among European operators, flight crew associations, aircraft manufacturers. From the existing groups in the context of the EOFDM, namely the Working Group A, Working Group B and Working Group C, the most suitable one could be the Working Group B, which mainly focuses on programming and equipment-related aspects, defining FDM algorithms, and identifying techniques for analysing flight data.

The working group could further elaborate on the development of standard definition cards for the set of precursors identified under current guidance material, further complementing them with particular triggering logic and further analytical considerations as it could be definition of severity levels or data pre-processing recommendations to support their standard implementation.

However, emphasis should also be placed on the difficulty that the working group may face in such a task. Many of the stakeholders involved are operators with mature and effective FDM programmes, in many cases tailored to their operational reality. In this sense, each operator has a deep understanding of its data silo, its actionability (data quality, availability...), as well as solutions for the resolution of constraints in its environment. However, in many cases, the standard solution implies finding trade-offs, which can only be defined when considering a multi-fleet and multi-operational environment. Indeed, the detail of the standards proposed in this solution can only be derived from participants' experience in processing the data. In such a collaborative group, it is likely to be necessary to establish a framework for sharing results and/or analytics, which allows working on the logics and conditions applied by each operator and the results obtained, guaranteeing the confidentiality and protection of the outcome and information at all times. In that regard, it should be noted that sharing results between operators is very challenging, because of the data protection principles applicable to FDM data (ORO.AOC.130). That is mainly because other operators may be competitors, and because many operators fear that their data may be misused by others. In addition, the intrinsic design of the FDM software of each operator may result in different results for the same FDM event definition. This can make comparisons of FDM events between operators using different FDM software challenging. Therefore, it is important to recognize the challenge involved in implementing this approach at the European level, beyond the successes achieved by similar initiatives such as the EOFDM.

The proposal contemplates taking collaboration a step further, including a sharing of pre-processed analytics results in the environments controlled by each data owner. In some ways, this could replicate results-sharing exercises similar to those that operators around the world conduct, in many cases on an ad-hoc basis, in trusted environments for the discussion of specific risk areas in regional or local working groups (e.g. CAST or RST). In many of these cases, the sharing of analytics is voluntary and the approach decided by each contributor. There are no previous agreements on the scope of data to be pre-processed nor on the pseudocodes used to calculate the results. Nonetheless, the sharing of results does exist, in contexts of trust under directed work groups which some MoUs or agreements.

Although it could be complex, governance mechanisms could be studied to share a certain level of analytics and an additional level of detail regarding the techniques and methods required for the identification of FDM events (e.g. high-level pseudocodes) without entering into areas that could infringe on the intellectual property of the providers (e.g. tailoring to fleets, operating environments, advanced techniques for data extraction...). Additionally, technologies and/or roles should be organised for the centralization of pre-processed analytics, with a technological complexity lower than that required for data sharing programs, but which could protect the recurrent analytical sharing exercise by data owners.

The following chapters provides additional ideas for possible solutions to the challenges posed by this approach, as well as its potential value primarily for international cooperation environments beyond the EU framework.

2.2.2 Approach #2 – Leveraging large data exchange programmes

Some States and organisations have set up so-called large data exchange programmes, under which they gather very large amounts of data (including FDM data) provided by many operators and by other industry stakeholders, which are then centrally processed and analysed.

Examples of such data exchange programmes are EASA's Data4Safety programme and FAA's ASIAS programme. Large data exchange programmes' participants also benefit from contributing their data because they are able to access the database and use the de-identified data for their own analysis and benchmarking. Operators may choose to share data from their FDM programmes among each other, for benchmarking purposes, although this needs to be done very carefully to obtain meaningful results. There needs to be sufficient similarity in place to allow any meaningful comparison between two operators. There is greater potential for achieving this, for example, between operators within the same group company, where normally there is both scope for data sharing and also commonality in terms of SOPs and operation of the same aircraft types. In short, the standardised definition and identification of FDM events is a key pillar in this type of initiatives, being another potential source for the development of this know-how across the different operators. This is the case of the example mentioned above in the development of guidance material for the identification of Unstable Approaches, elaborated by the EASA Data4Safety programme.

It is important to note that, although the development of standardised FDM events algorithms is a technical output of any data sharing programme, this does not necessarily imply that the programme enables mechanisms for sharing such technical specifications with its members, and therefore, the capitalisation of such knowledge. Additionally, it is important to highlight the origin of such standards, not necessarily being in all cases derived from the work of collaborative groups at industry level, but from the centralised expert work of the promoters of such initiatives. In this sense, the usefulness of this approach is to guarantee, on the one hand, that the harmonisation exercise results from the consensus of the industry, based on the capitalisation of common experience and, on the other hand, to ensure that this specific knowledge is disseminated and shared openly with the industry, finding the balance between the necessary information to support the advancement of the industry and the protection of the technological capital developed by the programmes.

In the European case, the recommended approach for the implementation of this mechanism is the EASA Data4Safety programme. In its Development Phase, the current methodology of the programme already involves testing and validating various algorithms and logics to identify safety occurrences from flight data, enabling benchmarking across operators and facilitating the sharing of best practices. Hence, the Data4Safety Programme provides a platform for enhancing the effectiveness of FDM programmes and improving safety performance at industry-wide level. This framework could facilitate the soft-level standardisation process by providing a platform for sharing validated approaches and promoting consistency across the industry. In that way, engaging with major industry players is possible, thus allowing smaller operators to follow the course. Moreover, although large operators may have definitions established by third-party FDM service providers, their participation in the Data4Safety Programme would help drive the adoption of standardised practices and/or promote understanding over their metrics and get them modified to better match to their operational reality. Hence, leveraging this existing programme could be highly effective due to its established infrastructure, stakeholder network, and focus on data-driven safety improvements.

2.2.3 Reflection on potential means to achieve widespread adoption of FDM events definitions

Achieving widespread use of standardized FDM event definitions presents a complex challenge, given the varied needs of operators and the flexibility required to adapt to evolving operational environments. The work done by international bodies initiatives such as the CICTT and EOFDM has laid a foundation for global standardization, but these efforts remain at a high level due to the inherent challenges of achieving alignment across diverse operators. These high-level frameworks provide valuable guidance, but they need to be translated into more actionable, operator-specific practices to ensure effective implementation and adaptation of FDM events to the operation (e.g. for the aircraft model, type of operation, SOPs, etc.).

The previously proposed solutions (Approach 1 and 2) aim to define the collaborative framework through which standards can be further developed to the level of details required for facilitating the adoption of the guidelines or recommendations by operators. Section 0 further explores how such a standard could be developed, using the reference event card from the EAFDM as baseline.

Regulatory changes, such as those proposed in NPA 2024-02, aim to introduce specific requirements to encourage broader adoption of these standards, but the full impact of these changes has yet to be seen. This raises an important question: what mechanism should be used to ensure the effective implementation of these standards? Should international standards like those proposed by CICTT FDMD WG and EOFDM be universally adopted, or should industry standardization bodies such as ARINC, EUROCAE, or SAE take the lead in developing a formal industry-wide standard? While this is a broad and complex topic beyond the scope of this document, a brief reflection on the matter is provided below, as per the information gathered during the DATAPP research.

The proposed changes to FDM regulation (NPA 2024-02)

The proposed changes in NPA 2024-02 on GMs and AMCs under ORO.AOC.130 emphasize the need for operators to maintain adequate knowledge of the flight data they collect and to fully understand the algorithms driving their FDM systems. This is crucial to ensure the correct interpretation and timely analysis of FDM outputs, which are fundamental to the SRM process. As per the proposed amendments, operators are required to have enough knowledge of their FDM algorithms to assess their adequacy for specific aircraft models, types of operations, and SOPs. This includes the ability to modify or request adaptations to predefined algorithms, which many operators still rely on without sufficient customization.

AMC1 focuses on ensuring that operators maintain oversight of the collected flight data and the FDM algorithms in use, emphasizing that the responsibility for the FDM program cannot be delegated. AMC2 supplements this by addressing agreement on key risk areas, encouraging operators to cover these areas with FDM. This is a vital step in aligning FDM practices with specific safety risks, but it remains somewhat high-level and leaves the door open for further development in operational customization.

Moreover, GM1 ORO.AOC.130 highlights the importance of adaptability in FDM event algorithms. The ability to adjust FDM algorithms as operational conditions change (e.g., new destinations, changes in SOPs) is critical for ensuring that FDM programmes continue to deliver relevant results. In this context, adjusting the variables in an algorithm is often more effective than creating entirely new event definitions.

At the time of development of this material, many operators still depend on predefined algorithms provided by their software or service providers, with little effort made to adapt these to their specific operational environments. The proposed changes are still under discussion and therefore, it is difficult to discern the effect they will have on the standardization of the events under monitoring in FDM programs, as well as on the adequacy of the events to the operational conditions of each operator.

In any case, as learned from exchanges with various stakeholders during the research project, many operators are already monitoring precursor events within the critical risk areas, often as part of the event portfolios provided by FDM service providers. However, the access to information regarding the algorithms used varies significantly between operators, with some having more more insight into the pseudocodes implemented than others. Most services provide the capability to adjust thresholds for certain parameters that trigger event conditions, though not all are supported by a comprehensive suite of analytical tools to help operators assess which thresholds best suit their operations. Likewise, the ad-hoc coding of metrics or their reprocessing in historical datasets is not always straightforward across all services.

Therefore, without a clear framework defining the dimensions that shape event definitions—those that should be controlled and understood by the operator—it remains unclear which aspects are intended for standardization (e.g., conditions triggering detection) and which are expected to be tailored (e.g., algorithm or method, detection thresholds, severity thresholds). It is necessary to establish a definition that enables the standardized adoption of safety events under the key risk areas, fostering a common understanding of these events. This would not only support the progressive integration of FDM program outputs into both national (SSP) and European safety ecosystems but also facilitate the identification of the parameters and conditions that need to be tailored to each operator's specific operations, along with clear guidelines on how to implement such customization. For this reason, while the proposed regulatory changes represent a step toward standardization and enhancing the effectiveness of FDM programs (e.g., through defining risk areas and monitored events, referencing the work of EOFDM), in terms of promoting widespread adoption, as per the opinion gathered through the research project, they may fall short in providing sufficient guidance for both authorities and operators to promote a more standardized framework for such events.

Returning to the question that closed the previous section: as we move towards defining a framework for industry-agreed FDM event definitions, what are the most effective vehicles to achieve this? Should we continue along the current path, or is it time to shift towards alternative solutions, such as industry-wide standards? The challenge lies in establishing a consensus that promotes standardization across operators while still allowing the flexibility needed to address varying operational requirements. This section will explore potential pathways, including regulatory initiatives, industry standards, and collaborative efforts, to identify the best approach for ensuring the widespread adoption and effective implementation of FDM event definitions.

A path forward: International standards, industry technical standards or regulatory solutions?

There are several potential paths to achieve widespread standardization while maintaining the necessary flexibility for individual operators. The key approaches can be broadly classified into three categories: international standards, industry technical standards, and regulatory solutions, each of which present unique advantages and challenges briefly presented below:

International standards via CICTT

International organizations such as the CICTT have laid important work for global standardisation of FDM events. These initiatives have helped at creating a shared understanding of key risk areas and safety events, fostering alignment across the industry. However and as previously mentioned, this approach comes with certain governance challenges in particular to establish an effective exchange of information at the level of technical detail needed to promote meaningful, in-depth discussions that drive actionable technical guidance. Initiatives focused on large-scale data exchange, such as Approach #2, have the potential to address this challenge.

However, the effectiveness of such efforts will largely depend on the representativeness and diversity of the participating organizations. Additionally, the absence of enforcement mechanism limits the practical implementation of these standards. The voluntary nature of adoption means that implementation might remain inconsistent across regions and operators.

Industry technical standards via ARINC / EUROCAE

Developing technical standards through industry bodies such as ARINC or EUROCAE could be an alternative solution. ARINC, SAE, and EUROCAE are well-established industry standardization bodies that have historically played a key role in developing technical standards for avionics and systems integration. These organizations have the capability and expertise to develop pseudocodes or technical standards for FDM event definitions, which could provide a structured, universally applicable framework for operators, manufacturers, and service providers.

These organizations could create precise and technically detailed standards, working closely with FDM software providers and operators to develop guidelines on event definitions, thresholds, and algorithms. The benefit of this solution would be the potential for innovation and technical precision. It could promote a higher level of customization, which is essential for operators with varied operational conditions.

However, their approach is typically more technical, focused on system interoperability, data formats, and equipment specifications. While this is useful, it may not directly address the more nuanced operational aspects of FDM customization and event definitions. The challenge with relying solely on these bodies for FDM event definitions lies in their scope and adoption process. Standards developed through ARINC or EUROCAE could provide a strong baseline, but widespread adoption would still depend on how these standards are implemented by airlines, regulators, and FDM service providers. The absence of enforcement mechanisms remains a challenge. These organizations have a formal and sometimes slow-moving standardization process, which might not be agile enough to keep up with the rapidly evolving needs of FDM and advanced analytics. Moreover, operators might hesitate to adopt these standards if they are seen as too rigid or not easily adaptable to individual operational requirements.

In addition, FDM event definitions go beyond just technical specifications. They require a deep understanding of operational contexts, safety risks, and how different data sources interact (e.g., flight crew reports, weather data, maintenance records, training…).

Regulatory solutions via AMC / GM

Finally, a more structured approach would involve regulatory solutions such as introducing an Acceptable Means of Compliance or Guidance Material, in line with the proposed changes in NPA 2024-02. This approach would provide a higher-level constraint by establishing clear regulatory expectations for FDM event definitions, in particular to further define the requirements in the newly proposed AMC1, GM1 and GM2 to NPA 2024-02. An AMC could specify baseline requirements to adopt effective FDM events as per standardized definitions, severity levels, and the parameters that operators should be allowed to tailor based on their operational needs without deviating from standard definition.

The introduction of an AMC would ensure consistent interpretation across operators and create more streamlined opportunities for safety cooperation at national and regional levels (e.g. contribution to data exchange and cooperative programmes)..

However, a significant issue with this approach is that it puts the burden of producing standard FDM event definitions on EASA. Given the complexity of the task, it would probably claim significant resources from EASA over several years.

Conclusion: Balancing standardization and flexibility through industry cooperation + regulatory framework

While each of these approaches offers viable solutions, with their pros and cons, the key challenge is reaching the right balance between standardization and flexibility, as FDM event detection methods and thresholds might vary between operators due to differences in fleet composition, operational environments, and risk tolerance. A flexible solution is crucial to allow operators to tailor their FDM algorithms to their specific needs while maintaining a standard framework that ensures consistency across the industry.

A common concern among the industry is to avoid a state of over-regulation, especially in an environment where many operators already have a mature FDM programme in place. Most operators share the need to promote the harmonisation of events and the benefits they would bring. However, they also stress the effort that would be required not only for their definition, but also for their implementation within a rigid regulatory framework, as well as the need for such flexibility in order to be able to capture the particularities of their operations. Furthermore, it is important to define under which parameters or dimensions such standardisation is established. Although consensus could be reached regarding the measurable conditions for triggering events, it has been argued that the detailed algorithm behind the event pseudocodes is part of the value proposition of FDM software vendors and that standardisation should be outlined in such a way that it does not undermine the efforts of these vendors to innovate and promote more advanced algorithms for effective identification of events or fostering quality of FDM data.

The potential challenges in establishing standardized severity levels have not been thoroughly discussed. However, researchers, drawing from their own experiences, believe that severity levels for events are often predefined by software vendors. In more advanced operators, these levels may be customized to align with their specific SRM processes. While standardization is feasible, the definition of a standard for event severity assessment should take into account the operators' responsibility to adapt their FDM programmes to their risk management systems. This includes tailoring severity assessment criteria in accordance with each operator's SMS.

Considering all this, the path forward should focus on a flexible, yet structured approach that ensures standardization where necessary while encouraging customization and innovation. An hybrid approach combining ongoing collaboration between industry and international bodies, as well as regulatory mechanisms could provide the right balance, ensuring that FDM event definitions are both effective and adaptable to the evolving needs of the aviation industry. Particularly, industry-agreed FDM events definitions could be further developed under one of the collaborative approaches presented above, potentially promoting a framework for greater data-driven cooperation for the assessment of more detailed practices for the proposal of such standards. This would include development of 'reference event definition' cards, including standardized triggering conditions for events, parameters required, severity levels and high-level pseudocodes, as well as parameters to be tailored as per the operation. At the end of the process, an introduction of such reference event definition into AMC could be the most effective way forward to achieve adoption of such standards. It would create a regulatory baseline for FDM events, ensuring a minimum standard for their identification while allowing operators to tailor aspects such as detection thresholds. This approach would not only support standardization but also provide the necessary room for innovation and differentiation.

2.3 General recommendations for definition of standardised FDM events

A standardised FDM event must be versatile enough to apply to all types of operations and airplane models, making a single, fixed algorithm impractical for its trigger logic. However, achieving some level of standardisation at 'pseudocode level' is important for facilitating data collection and comparison among different operators, allowing for meaningful analysis and aggregation. Similar to the recommendations issued by the EAFDM, a flexible definition is to be developed.

Under this flexible approach, the trigger logic remains consistent, but the values of the variables within the logic can be adjusted by operators based on specific aircraft models, airports, or standard operating procedures. When applied under a collaborative analytical environment, a certain level of standard should also be promoted at aircraft model & airport level. Instead of using hard-coded threshold values, the indicators refer to aviation operational concepts and provide a range of acceptable values (e.g., CAS thresholds between 50 kts and 80 kts).

Similarly, while guidance is provided for indicating severity, fixed threshold values should not always be strictly defined to allow for customisation. The goal is to make the standardised FDM event easy to understand and implement. Thus, the definition should be straightforward and clear, use flight parameters that are commonly recorded on most airplanes, and feature trigger logic that can be easily programmed using logic and arithmetic operators available in most FDM software. If possible, a mapping of flight parameters naming across different data frame layouts should be included, supporting implementation across different fleets.

It is also important to recognise that implementing standardised FDM events is an iterative process. Operators may need to fine-tune these indicators over time to ensure they are effective and provide the desired outcomes.

2.3.1 Reference event definition card

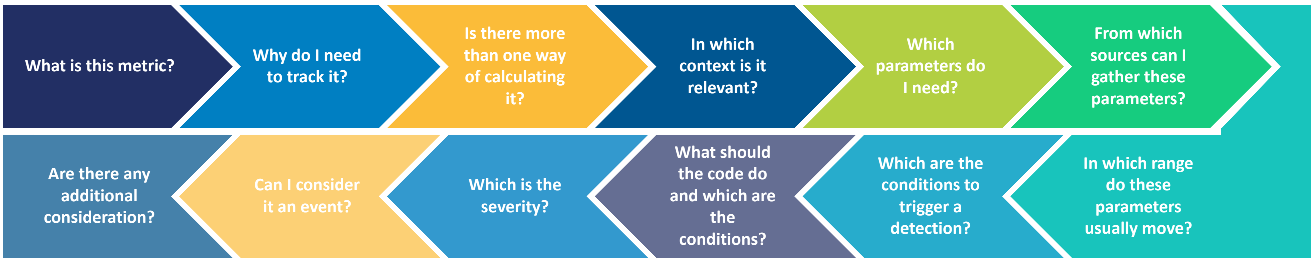
As a key step towards standardisation, it would be recommendable to develop a standardised definition card that includes an agreed definition that could be implemented by any stakeholder. As previously explained, a reference "FDM-based indicator" card is provided in the context of the work of the EAFDM, for each of the standardised FDM indicators. Such work was also used by FAA's ASIAS programme for the definitions of their FDM events.

The provided EAFDM card includes the following fields:

- **Title:** title or identifier of the standardised FDM-based indicator.
- **Description:** short text describing what potentially unsafe situation the indicator is designed to detect.
- **Applicable operational risk:** risk information that the standardised FDM-based indicator is meant to capture and the occurrence category for which it is relevant (e.g., Runway excursion).
- **Trigger logic:** combination of conditions that triggers the detection and the list of data needed (e.g., flight parameters).
- **Suggested indications of severity**: recommended indicators for assessing the severity of the situation.
- **Useful contextual information:** Information not captured by FDM data which is useful for the analysis of the trigger logic detection.
- **Effectiveness:** considerations on the limitations of the trigger logic.

This card is proposed to be used as a baseline, to be further complemented with some additional fields to provide further guidance to the main stakeholders when implementing the standardised event definitions. More specifically, the cards should contain all the fields which could help to give an answer to the following questions:

Figure 2-1 Questions to be answered by the reference event definition card



Considering such questions and the fields added in the other works and referenced material, some recommended additional fields could be:

- **Rationale** for monitoring the metric
- Potential **standardised variants of the metric**, providing some flexibility to the different stakeholders' contexts
- **Metric scope or applicability**, indicating the context or scenario in which applies

- **Minimum set of flight parameters**, defined in a separate field
- **Data sources** from which to collect the different needed parameters
- **Operational ranges for parameters,** indicating the upper and lower limits of the values at which they usually operate
- **Conditions triggering a detection**, similar to what is provided in the trigger logic field of the EAFDM cards, but only providing the high-level conditions (measurements) to be evaluated for determining if a detection is triggered, using a separate field (e.g., high airspeed)
- **Algorithm or method**, providing the full detail of the step-by-step process that the code follows and the criteria related to the different parameters and their defined thresholds which allow to evaluate if a detection is triggered
- **Severity level assignation**, providing the method to be followed to define different levels of severity and to subsequently assign an established severity
- **Event identification**, establishing the conditions to consider an event based on defined criteria
- **Additional considerations** that should be taken into account in the algorithm or the logic

Consolidating the aforementioned fields, the following card structure could be used as a reference. Such card would reflect the different key elements to be defined for each of the standardised FDM events.

Table 2-1 Reference event definition card

FDM Event:	**Name of the FDM event**
Precursor metric name:	**Name of the precursor metric**
Metric definition	A clear and concise statement explaining what the metric measures in the context of flight operations.
Rationale	The underlying reason for monitoring this metric, including its importance for safety, efficiency, or regulatory compliance.
Variants	Different forms or versions of the metric, if applicable, that may be used by the stakeholders for various types of analysis or comparisons.
Metric scope / applicability	The specific contexts, phases of flight, and types of aircraft to which the metric is relevant.
Minimum set of flight parameters needed	The essential flight data parameters required to accurately calculate and assess the different established criteria.
Data sources	The primary and secondary sources from which the necessary data is obtained, such Flight Data Monitoring programmes' data, weather data or specific aviation related datasets.
Operational ranges for parameters	The defined normal ranges for the parameters that are monitored, indicating acceptable conditions during operations.
Conditions triggering a detection	Specific criteria that, when met, indicate an event or anomaly that warrants further attention or action.

FDM Event:	Name of the FDM event
Precursor metric name:	**Name of the precursor metric**
Algorithm or method / “trigger logic”	The computational approach or process used to derive the metric from the flight data or other specific data sources, including the specific thresholds for each considered criteria, and any formulas or algorithms applied.
Severity level assignation	The method for categorising the severity of each triggered or met criterion, including definitions of different severity levels and their considerations and implications.
Event identification	Conditions for which an event is identified, for instance considering the number of met criteria and the severity level assigned to each of them.
Additional considerations	Any additional information relevant to the metric, such as limitations, assumptions, potential improvements, or special conditions affecting its use.

An example of the table filled in with the information for the case regarding unstable approach can be found in section “Annex A. Illustrative example of FDM event definition card - Unstable Approach ” of this same document.

2.3.2 Flight data analysis techniques and principles for standardised FDM events

This sub-chapter aims to provide a brief guide to the analytical techniques and principles to be followed for the definition and implementation of standardised FDM events, based on the consultant’s experience and considering the recommendations provided in the “FDM analysis techniques and principles” document [4]. The output captured on the cards is a simple synthesis of the analytical work required to derive standardised metrics. As discussed above, the definition of such metrics requires a factual discussion, based on analytical exercises where various stakeholders share the results and/or, ideally, the data from their FDM programme to be in a better position to define standardised metrics.

However, the process followed for the definition of such metrics will largely depend on the governance framework of the collaboration between the stakeholders involved. For example, within a working group, the governance framework may provide a forum for the discussion of pseudo-codes, technical specs and results obtained from the application of conditions on the data of the operators involved, without being able to analyse the data as a whole. In the framework of data sharing programmes, it is possible to go a step further, retaining control and ensuring real harmonisation in the implementation of pre-processing steps and triggering conditions, as well as in the analysis of outcomes thanks to the common analytical platform.

The following is the set of steps that should be followed for the definition and implementation of a pre-selected standardised FDM event in the environment of maximum data availability. However, considerations are provided for those scenarios where the confidentiality framework requires greater restrictions.

Discussion on the rationale, metric scope & applicability

The first point of discussion is to encourage a discussion on the rationale, background, scope and applicability of the safety event to be defined. This discussion should take place within the collaborative governance framework of the programme or group leading the initiative. In the context of this discussion, a common understanding of the specific definition of the event to be identified is key.

As an example, it is important to map the event to the reference taxonomy, as FDM event groups identified in the AMC1 ORO.AOC.130, the applicable operational risk or Key Risk Area identified within EASA Annual Safety Review, ECCAIRS taxonomy or occurrence categories, as the CICTT. This is especially important in an environment where several operators work with different FDM software providers and/or different fleets, both sources of disparity in the nomenclature given to events. The initial discussion should also lay the groundwork for the analytical cooperation plan, focusing on the development of common metrics, which might involve the following topics:

In the case of collaborative initiatives that do not involve a common repository of shared data:

- Agreement on the sharing of pseudo-codes and/or technical reference material for the identification of events by operator.
- Agreement on the scope of data / reference scenario / use case for: 1) testing of the harmonised algorithms (in case it is part of the scope definition; 2) Sharing of results of metrics implemented in each operator, to study potential differential points in definitions and/or pre-processing, to derive common lessons learned.
- If not defined at working group level, the conditions of de-identification and confidentiality in the sharing of the results of metrics and pseudo-codes.

Additionally, in case of initiatives with common repositories for the analytical exploitation of shared data:

- Agreement on the shared sample of data for implementation of the standardised metric.
- Agreement on incremental implementation, validation and testing of standardised algorithms.
- Validation plan by comparing results of harmonised algorithm versus own results of participating operators, to derive lessons learnt and iterate on the improvement of the proposed definition.

Discussion & application of general pre-processing rules

The first point of discussion and/or the initiative should focus on the minimum data quality conditions to enable the implementation of the proposed events. In the case of initiatives that do not involve data sharing per se, this point is of less relevance, being relegated to the guarantee that each operator participating in the exercise provides minimum guarantees that its technological environment ensures data quality metrics. In the case of environments in which data are shared, this point is crucial, but it is assumed to be a fundamental task in the development of the pipelines necessary to make the data from such programmes available.

Figure 2-2 Main data quality checks related dimensions

Overall, a set of pre-processing rules must be applied at a general level for the collected flight data set. These pre-processing rules should mainly be applied to a defined scope of parameters, prioritising the most commonly used ones (e.g., mandatorily-required parameters or most commonly used ones). The aim of the application of such rules is to perform flight data quality checks for ensuring that the following dimensions, which are detailed in the EOFDM document "FDM analysis techniques and principles" [4], are properly addressed:

- **Completeness:** Proportion of complete data ingested (checking data gaps in flights / flight parameters).
- **Validity:** Data are valid if they conform to the syntax (e.g., check outliers, operational range for continuous parameters & discrete values for discrete parameters).
- **Accuracy:** Data correctly describes 'real world' situation (e.g., bias correction algorithm for normal acceleration). Ad-hoc correction of parameters should be applied at metric-based level, if required.
- **Consistency:** Values and variations of flight parameters are consistent (e.g., cross-check correlation exercise).

- **Uniqueness:** No elements are recorded more than once (e.g., duplication of flights / flight parameters).
- **Timeliness:** Recording rates are sufficient for the application of the measurements. Minimum recording rates per parameter will be discussed and filters will be applied accordingly.

Performing these quality checks through the implementation of data quality algorithms enables to address the typical data quality issues that can arise and that are described in the following table:

Table 2-2 Main data quality issues to be addressed by the data quality checks

Data quality issue	Description
Data gaps	Data gaps refer to missing data points in the recorded dataset. This can happen due to intermittent sensor failures, data transmission issues, or recording interruptions. Flight parameters are not recorded properly at specific flight phases or recurrently.
Spikes above operational ranges	Sudden, erratic and not prolonged deviations in the data that exceed the normal operational range of the parameter. They often appear as sharp peaks in the data.
Toggling	Rapid and frequent switching between two or more values. This can occur due to issues such as electrical noise, poor sensor calibration, or mechanical vibrations affecting the sensor.
Frozen values	At a given point of a flight, a parameter 'freezes' and it is no longer properly recorded (due to malfunction or parameter behaviour). Data points remain constant over a period when they are expected to change. This could be due to a sensor getting stuck or a failure in the data acquisition system.
Offsets	There is a constant difference between the flight parameter value and the actual physical value. This may occur to flight parameters that require calibration (e.g., normal acceleration).

As discussed above, either for programmes where these rules must be implemented and/or where results are only shared for the purpose of defining common metrics, specific metrics should be agreed and monitored around each rule to validate that the quality issues are addressed (e.g., checking the completeness by monitoring that the cuts (missing value %) in flights that are imported represent less than 10% of the total duration of those flights).

Then, a potential approach is to tag the flights according to the identification of issues based on the proposed pre-processing rules. Thus, those flights that do not fulfil the defined requirements or conditions can be identified and quarantined, hence not being propagated to the metrics calculation phase. Moreover, the share (%) of flights per aircraft type/configuration that are left out of the processing due to the rules, before and after their application, should be monitored. In the case of programmes where each stakeholder with analytical capabilities analyses his own data, this may affect the representativeness of the pre-selected use case. In such a case, once the exercise has been carried out by all operators, a sharing of such indicators may favour a discussion where it is decided whether the use case selected to start the definition of the metric is representative and reliable, at the data level, or it is necessary to change it.

Quality assurance for flight phase splitting

A further step is to ensure that flight and flight phase splitting is adequate between participating operators. Similarly, quality assurance rules should be applied to them to ensure that the necessary requirements are met. This is usually a basic functionality in the context of any FDM software, but it is possible that the logics implemented for flight phase separation involve differences that, although not very relevant for most use cases, could be sensitive for those events that are identified in the phase transition or require high precision (e.g., precise identification of the touchdown milestone on the runway).

In the case of large data-exchange programmes, although the availability of data for different dataframes should be considered, the objective should be to standardise the definition as much as possible. By uniformly defining the various phases of flight such as taxi, take-off, climb, cruise, descent, and landing, the algorithms ensure that the data is segmented in a comparable manner and it allows for more accurate identification of trends and patterns in flight operations, facilitating the detection of safety issues specific to certain flight phases.

Drafting of the pseudocode, exploratory analysis & selection of key FDM parameters

At this point, having ensured the quality and actionability of the data, a draft of the event logic should begin to be developed. For this purpose, a critical discussion of the pseudo-codes implemented by different operators, representative of different fleets and dataframe layouts, would be favourable. This may be a potential limitation, given the limitations identified in the context of the DATAPP project, regarding the capitalisation of knowledge on the logics of some operators, due to multiple reasons, among others, the opacity of certain FDM analytics solutions to such information. Confidentiality agreements in the working group are relevant to develop a trust environment that favours the sharing of such information. In addition, the participation of third parties (non-operating stakeholders) could facilitate the intermediary role for the development of this exercise of comparison between pseudocodes.

The exercise would entail identifying the basic conditions and measurements for the identification of the analysis window, the event and its safety classification. It is of special interest to identify the required data sources and parameters, covering different fleets and data frame layouts. A simplified and harmonised condition could be derived from this exercise, acting as a sort of high-level pseudocode and a master list of minimum required parameters. All necessary parameters should be identified and mapped to the different dataframes available, since a parameter may have different names or values for each dataframe.

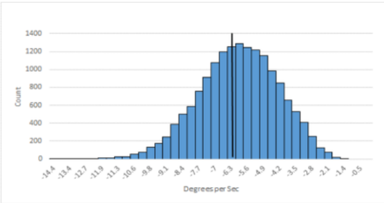
Table 2-3 Example of the table for mapping the parameters across dataframes

Flight Parameter	Parameter in Dataframe 1	Parameter in Dataframe 2
Flap_pos	FLAPS	FLAP1
Land_gear	LDG_NOSE	NLG_SYM

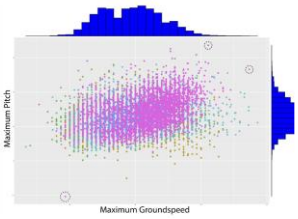
As explained in section “Reference event definition card”, the standardised event definition cards would provide information on the parameters and the algorithm to be used. Consequently, there is further work needed to ensure that such minimum required parameters have the appropriate quality and standardisation to be considered by the defined algorithms. It is essential to confirm and understand the relevant aircraft flight parameters’ performance before commencing the process of programming and implementing any such standardised FDM-based indicators. Only valid, accurate and correctly sampled flight parameters can deliver a meaningful indicator. In addition, some of the proposed standardised FDM-based indicators may not be achievable, because the needed flight parameters are not recorded with sufficient sampling rate, accuracy or recording resolution. It is acknowledged that less advanced or similar indicators may be implemented instead.

Considering that, an exploratory analysis could be proposed to understand the applicability and standardisation of each parameter for specific events, evaluating the operational range (maximum and minimum values of the parameters), sampling rate, completeness, and maximum rate of changes of preselected parameters, for each aircraft type and, ideally, per flight phase. Such exploratory analysis is recommended to be supported by key visualisations such as histograms, correlation analysis or box-plot analysis. As an illustrative example, the following examples of key visualisations which are provided in the “FDM analysis techniques and principles” document [4] developed by EOFDM could be used:

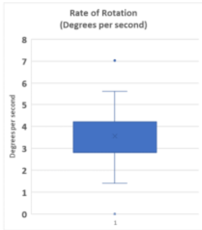
- **Figure 2-3** Key visualisation types to be considered for the exploratory analysis. Source: EOFDM WGC - “FDM analysis techniques and principles”



Histograms



Correlation analysis



Box-plot anaysis

Although in the best case, a centralised shared data environment would facilitate this exploratory analysis, it could also be performed as an analytical exercise by each contributing operator in their own silos. The performed exploratory analysis of each parameter would allow to develop a table including all the relevant information to be considered for the event definition, which would complement the already provided information on the cards such as the list of parameters itself, the type (continuous or discrete), the data source, the operational ranges and if it is mandatory or not:

- **Table 2-4** Example of the structure of the results from the exploratory analysis

A/C type	**Flight phase**	**Flight parameter**	**Completeness** (% of missing values)	**Operational range** (max – min)	**Median**	**Q1**	**Q3**	**Avg. sampling rate**	**Max. rate of change**	**Min. rate of change**
AT1	Climb phase	Fuel mass flow rate SYS 1	95%	2500 - 1700	1900	1750	2200	4	900	-500
AT1	Cruise	Fuel mass flow rate SYS 1	…	…	…	…	…	…	…	…

The results of the exploratory analysis on the pre-selection of minimum parameters should facilitate the decision on the suitability of such parameters for the standardised metric. In other words, identify which minimum parameters are necessary to apply the common logic to a set of representative dataframes and fleets.

Application of specific data quality rules at parameter level

Despite having a minimum parameter environment to use, this does not guarantee data quality in any flight and flight phase. Multiple conditions can affect the quality of the data and it is therefore necessary to define a minimum of quality rules that must also be reflected in the reference definition card. Therefore, the definition of quality rules is recommended, this time at parameter level, to tailor the parameters to be used for the identification of specific events, thus ensuring the right accuracy, completeness, and validity of the different parameters.

In the case of analytics shared between individual operators, such information should again be provided by the operators, based on the pre-processing rules they are applying and/or their FDM provider, in case of outsourcing. In the case of collaborative data programmes, parameter-level quality is the analytical step that follows an exploratory analysis. As an illustrative result, it would be extremely useful to have a table that indicates the key quality assurance rules to be applied to each of the parameters, as well as any data preparation considerations. The following table contains a non-exhaustive example of the set of rules that could be considered for some specific parameters:

Table 2-5 Non-exhaustive example of the parameter-level quality assurance rules to be applied

ID	Data source	Aircraft parameter	Data preparation considerations	Operational ranges check (static THR)	Dynamic changes check (rate THR)	Offset / Bias correction	Interpolation (increase sampling rate)	Smooth time series
FM.3	Flight data	Ground speed	…	Required	Required	Not required	Not required	Not required
FM.4	Flight data	Indicated airspeed	…	Required	Required	Not required	Not required	Not required
FM.5	Flight data	Pressure altitude	Dependent on operational / atmospheric conditions and for certain applications, contextual data needed for correction.	Required	Not required	Not required	Potentially required	Not required
FM.6	Flight data	LAT/LONG	…	Required	Not required	Potentially required	Potentially required	Not required
FM.7	Flight data	Time	…	Required	Not required	Not required	Not required	Not required
FM.8	Airport data	Airport location	…	Not required	Not required	Not required	Not required	Not required

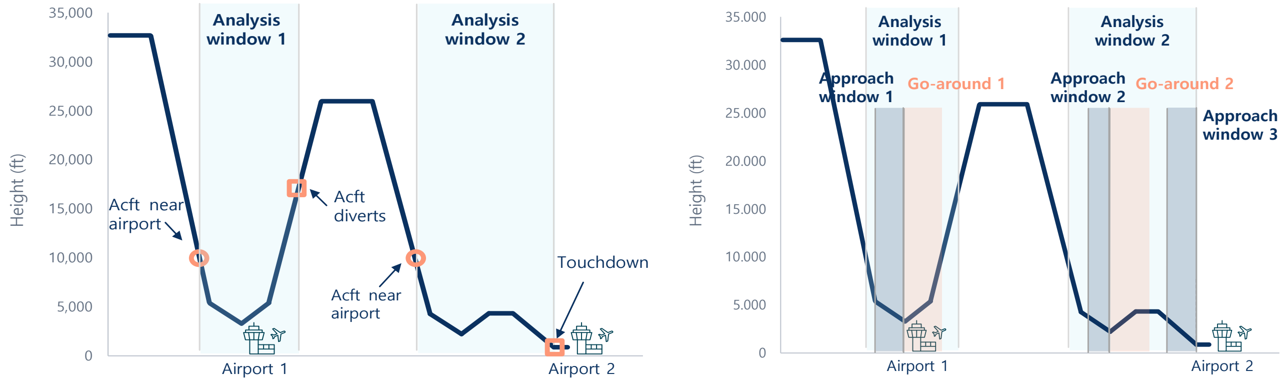
This type of table containing the requirements in terms of the main data quality checks to be performed over each of the parameters could be provided as further guidance, additionally contributing to standardising the parameters used and, thus, the event definitions. By providing this combination of potentially standardised data quality procedures together with the provision of the standardised event definitions cards including all the relevant fields, the stakeholders would have clear guidance to implement a standardised framework for identifying specific events.

Definition of the analysis window

The 'analysis' or 'search' window aims at defining a subset of flight data to be processed for the purpose of the identification of events. It aims at rationalising the amount of data to be processed by the more complex algorithms and commonly relies on flight phase splitting logics.

In this sense, as mentioned above, although similar, flight phase splitting logics may differ between different providers, so that the analysis windows may be affected. At this point, validations can be introduced with respect to the identification of the analysis windows, based on the number of windows identified per flight, the average duration of the windows and the distance travelled during the windows.

Figure 2-4 Identification of the analysis window



Definition of the trigger logic of the algorithm

The trigger logic definition entails grounding the high-level pseudocode discussed at early stages of the initiative, on the basis of the exploratory analysis conducted. Such analysis allows to determine which is the normal situation or scenario complementing the operator's specific documentation. Specifically, the study of data distributions by parameter and/or the statistical metrics mentioned above, are relevant for the definition of triggering conditions. Each operator may have implemented specific thresholds and a high-level pseudocode, on which to base further discussions to be conducted. In that regard, the objective is to establish an indicative range, whether than specific thresholds, for the selected use case and understand which are the key drivers for establishing it at fleet or SOP level. Again, there are differences between the potential approaches that could be followed. In the case of the data exchange programmes, specific values should be defined as technical implementation of the algorithms is expected, and cooperative analysis of results is commonly part of the scope of such initiatives. However, regarding the collaborative programmes in which there is not a sharing of data, the aim is more focused in fostering discussions on that domain.

Considering the standardised event definition cards, the trigger logic would be a field to be provided, since it clearly marks the level of standardisation as it defines the specific conditions. The trigger logic may be based on several conditions, or on the fulfilment of one condition over a sustained and defined time period (e.g., 5 seconds).

The purpose is to draft the key conditions to be coded for the algorithm to identify the event, which should include:

- **Triggering conditions:** One or more FDM measurements calculated on specific parameters, the fulfilment of which captures a state that triggers event identification.
- **Confirmation time:** Minimum time duration where one or more conditions must persist to confirm the identification of the event, making the identification more robust to value swings at specific timestamps.
- **Thresholds / Severity levels:** Specific values or range of values above / below / in which the identified conditions are triggered. These may additionally imply different categorisations of events with respect to their severity (see next section).

Definition of thresholds and severity levels for FDM events

FDM events should be classified according to their severity, meaning the score allocated to an FDM event according to pre-defined criteria that is determined based on FDM data. It is important to note that the severity score assigned to the event differs from the risk level of the event under analysis, which is derived from the application of risk analysis methodologies. Nevertheless, the severity score can be a valuable metric to feed into the risk assessment process.

Ultimately, consistency in the application of severities to all algorithms that determine FDM events is key to meaningful indicator analysis at operator SMS level. Additionally, in the aggregation of risk indicators at the national or international level, consistency in the application of severities at the FDM event level can lead to harmonisation of the data collected and, potentially, to better profiling of the risk at industry level. The standardisation of severity levels goes hand in hand with the discussion of thresholds, as well as the definition of a framework of 'leading' and 'lagging' indicators to identify reductions in safety margins and excesses, respectively.

The assignment of severity scores also falls under the framework of flexibility and adaptation of the FDM programme to the operational reality of each SMS. Therefore, in most standardisation initiatives, severity is proposed as a set of ranges on certain conditions or metrics, as a guide for operators to progressively harmonise their data-driven method of assessing events and thus facilitate more harmonised discussions. In common analytical frameworks, under data-sharing programmes, specific thresholds should be set to feed into risk indicators.

As mentioned above, it is necessary to establish a common framework for defining severity levels. As an example, the work carried out by the EOFDM, specifically in the “FDM analysis techniques and principles” document [4], proposes the standard definition of four levels of severity score:

- **Level 0:** Level 0 does not necessarily constitute a safety hazard, but more momentary or limited exceedances of measurements, that might indicate deviations SOPs, among others.
- **Level 1:** Level 1 events may occur frequently, capturing an approach or exceedance of a primary barrier. Monitored at aggregated level, they measure a reduction on the safety margins.
- **Level 2:** Level 2 are indicative of actual safety margins reduced to the final safety barrier. They are rare and require a consequent root cause analysis.
- **Level 3:** Level 3 capture the undesired state per se. A full operator (and possibly State) investigation would typically be performed on each individual Level 3 event.

At this point, the difficulty lies in defining ranges of values for each of the severity levels. In other words, to define what the ‘nominal’ operational domain means. To this end, multiple approaches can be followed, the feasibility of which will depend on the collaboration and data sharing framework of the group leading the initiative. For this discussion it is of relevance to have the definition of severity levels for the event, rationale and thresholds applied by the different operators, and ideally, access to the results of such metrics, categorised by severity (e.g., event rate by severity score). To propose the thresholds, two fundamental ways are proposed:

- **Knowledge-based / Procedural-based / airworthiness-based thresholds:** For certain types of events, the identification of leading or lagging indicators is pre-defined by business rules, either for airworthiness-based reasons, manufacturers' recommendations and/or procedural reasons. In these cases, the complexity of standardisation lies in mapping all issued recommendations to different granularities of analysis (e.g., SOP, aircraft type...), as well as reaching a consensus on which conditions must be fulfilled for triggering the event as such. As an example, a TAWS 'sink rate' alert could be considered by consensus as a Level 1 severity, although each operator may consider a minimum duration of the condition and/or other additional conditions to assign such severity.
- **Data-driven thresholds** (outlier detection)**:** In certain cases, it is possible to establish the event limit based on statistical methods on the measurements or conditions that compose the event. It is of interest, for example, for the determination of level 0 and level 1 metrics, indicative of exceedances that occur frequently but reduce safety margins. By studying the statistical distributions on different metrics, it is possible to establish ranges of values that can serve as triggers to measure these reductions in safety margins. As an example: In monitoring vertical speed as a condition that captures

instability within the Unstable Approach event, it is possible to define level 1 severity by evaluating the population of vertical speed measurements over the analysis windows of a representative population of flights. The threshold for level 1 could be defined for the 85% percentile of vertical speed values identified, and progressive thresholds could be promoted to study the tails of the histograms. Similar statistical methods could be introduced, as using interquartile ranges for lower / upper boundaries definition at different granularities. Multiple Machine-Learning based methods for multi-variate scenarios might be considered (e.g., Euclidian distance-based clustering techniques), which are particularly interesting for determining thresholds in those events where several parameters / conditions are evaluated. The data-driven approach is more feasible in those environments where data sharing and centralised analytics are favoured, being more difficult to implement if the analytics are produced in silos, as representativeness of data under analysis might be more limited. Note that thresholds driven by data-based methods need for a more robust review mechanism to reassure that they stay relevant for the operation along time.

As a general recommendation, the thresholds proposed should be more conservative than the thresholds determined by aircraft manufacturers or by operators' SOPs, in particular for leading indicators, so that the proposed algorithms provide meaningful indicators to track trends at operator and/or regional FDM level.

Figure 2-5 Definition and classification of severity levels

Testing and validation of the algorithms

The proposed solution worked out in this document aims to take a step further towards the standardisation of analytical methods and/or algorithms used by operators in the implementation of their FDM programmes. Without undermining the designed flexibility, it is considered important to provide guidelines of greater technical depth to support less mature operators towards the capitalisation of analytical knowledge, and with it, the improvement of their FDM programmes. For this reason, we consider the testing and validation phase of the algorithms as a fundamental step in this type of initiative.

We understand that in collaborative working environments without data sharing, as it is the case of the EOFDM / EAFDM WGs, discussions are mostly theoretical, based on the exchange of information and experiences between the participating experts according to the know-how of each operator. We recognise the complexity of carrying out more technical work in such environments, given the effort involved and the lack of a sufficient technological, analytical and data governance to favour such exchanges. In large data-sharing programme environments, however, this task is critical to the completion of the technical work assigned to the analytical working group and the generation of consensus intelligence. It is the cornerstone of the trust on collaborative analytics and a required step of the formalised approval process for technical developments. In addition to these technology related limitations, as has been previously mentioned, there is also the data protection & governance component, since operators may be competitors and because of the fear towards potential data misuse or leakage.

In the application of the recommended algorithms, operators may experience false positives due to unidentified data quality rules or a low/high number of events, due to non-fine-tuned thresholds. False negatives can be partially verified through cross-checking with reports. Regardless of the environment, we believe that such working groups should favour a validation of the recommended algorithms, commensurate with the governance of each working group.

In the following, some key points to cover under this activity are discussed, giving considerations according to the governance framework of the collaborative initiative:

How to build a testing plan for industry-agreed FDM algorithms?

Some examples are provided below on the key steps to implement a testing plan over the standardised FDM events algorithms. It covers two scenarios: large-exchange data programmes with centralised data platforms, and individual operator conducting independent analyses.

- **Scenario 1. Testing plan for large-exchange data programmes:** In a centralised data platform where flight data and reports are integrated at scale, a trusted third party performs comprehensive analysis and validation under an agreed governance framework for all integrated flight data and other data sources, if available. An illustrative testing plan would include:
    1. First, the analytical team evaluates the FDM event algorithm using recordings from flights with known events (e.g., reports from fused occurrence reporting system, if available at scale). It must be verified that the algorithm triggers accurately for most of these recordings, ensuring a low false negative rate or a low rate of missed FDM events. It is important to note that this validation, although useful to have a generalized view on the most relevant consequential events, is not usually entirely useful in the validation of precursor events and/or incidents. There are several factors that come into play, from the reporting culture of each organisation to the disparity in the perception and/or understanding of the events. For example, an unstable approach identified by the algorithm may not have been perceived by the pilot in command, spurious TCAS warnings recorded in the FDM system may not have been signalled in the cockpit due to the differentiation of systems and different logics in the onboard systems...
    2. Then, the FDM event algorithm is applied to a sample of flights over a defined period, such as a week or a month. The rate of nuisance FDM events is estimated within this timeframe, as this rate directly impacts the workload for manually validating FDM events. The resulting rates are compared against results provided by operators and/or other organisations. This comparison allows for a more robust and generalized estimate of the rate of events and identify outliers at operator / aircraft type level, which might lead to additional analytical work to understand if they are derived from real operational drivers or there might be a need for thresholds and/or pseudocode adjustments.
    3. Tailored validation exercises with operators can be launched based on their specific results. This approach ensures that each operator's unique operational context is considered, leading to more precise and relevant validation outcomes.

4. At last, the testing activity should gradually increase the sample size and include a diverse range of aircraft models within the scope of the FDM event algorithm. This incremental approach helps build confidence in the algorithm's trigger logic. The large, integrated dataset enables the establishment of reliable statistics on nuisance FDM events, facilitating continuous improvement and refinement of the algorithm.

- **Scenario 2: Testing by each individual operator:** In this scenario, individual operators conduct testing within their systems and derive insights to validate the proposed industry-agreed metric. It is assumed that operators have analytical capabilities in their FDM environment, including the testing of new algorithms on historical samples. As contrasted in the context of the DATAPP project, this is highly variable from vendor to vendor and in some cases, analytics on historical samples is reasonably limited.
    1. Each operator tests the FDM event algorithm against recordings from their own flights with known events in their own FDM software / platform, such as those reported in their occurrence reporting system. Additionally, operators can compare results with those from previously implemented algorithms using confusion matrices. This comparison helps identify true positives, false positives, true negatives, and false negatives, providing a clear understanding of the new algorithm's performance.
    2. Operators then can apply the FDM event algorithm to a sample of their flights over a set period (e.g., a week or a month) to estimate the rate of nuisance FDM events. Once again, the rate impacts the operator's workload for manual validation. Each operator analyzes their data independently, deriving insights that reflect their unique operational environments.
    3. After that, operators can progressively increase the sample size and include various aircraft models within their testing scope. This method allows them to build confidence in the algorithm's trigger logic gradually.
    4. Operators synthesize their analyses and derive lessons learned, which are then shared with other operators in a de-identified manner. This can be facilitated by a trusted third party responsible for aggregating and disseminating all results. This collaborative approach ensures that insights and improvements benefit the entire industry while maintaining confidentiality.

Data & technological framework required

As mentioned above, it is recommended that at the beginning of the work a data perimeter is established under which the discussion of the algorithms will be favoured. Although potentially expandable, depending on the event to be implemented, this favours a rationalisation in the analytical work of the previous activities, as well as in the validation exercise. Note that, in any case, the starting premise is a common technological environment among the participants, for the exploitation of historical flight data.

In this sense, a common operational/geographical environment should be decided upon among the participants, as well as a common time frame. This will facilitate the fulfilment of the minimum representativeness requirements for the sharing of results if this is possible under the initiative governance. The strategy for the selection of the data perimeter varies widely, depending on the event to be validated and the governance framework of the collaborative initiative and its requirements in regard with data sharing and de-identification rules, but here are some non-exhaustive recommendations:

- **Common operating environment with sufficient representativeness:** It is recommended to establish a common operational perimeter, on which the different operators and/or contributing organisations can discuss the particularities that may facilitate higher or lower rates of event identification, as well as discussions on the triggering levels of each condition. In turn, it is important that such an operational environment (airport / FIR / TMA...), allows sufficient representativeness to guarantee the rules of data protection but as well deriving reliable results. Although it is not the purpose of this paper to discuss methods for de-identification of results in collaborative analytical environments, a possible approach to determine the data perimeter would be to ensure:

  - At least, three different operators are contributing with flight data in the operational context defined (geographic scope + temporal scope)
  - A minimum number of flights might be considered from each operator contributing to the specific subset of data, to be counted for rule-of-three purposes and ensuring minimum representativeness for reliable calculation of rates.

- **Ensure aircraft fleet representativeness:** Beyond guaranteeing the minimum number of flights, it is important to ensure the representativeness of the fleet, especially in those algorithms highly dependent on conditions based on parameters or conditions highly dependent on the aircraft model.
- **Considerations regarding operational reality and/or seasonality:** Certain events may be more likely to be identified in certain seasons, either due to higher levels of congestion and/or other operational conditions (e.g. weather). It is important to take this into account, as well as the implications it may have on the representativeness of the results and the de-identification rules mentioned above. As an example, in certain seasons it is possible that more unstable approaches due to bad weather could be identified in certain destinations but that, at the same time, these are low seasons where the representativeness of operators and/or fleet is limited.

Again, it is important to note the complexity of carrying out such a validation exercise by each individual operator or when exchanged data samples are limited. In terms of representativeness, for example, it is difficult to assess compliance with the minimum requirements for sharing results if analysis is performed by operators alone. In addition, if the fleet size is small or for the selected data perimeter the flights are performed at irregular time intervals, it may take longer to build a suitable data set to test the FDM algorithm. When aircraft usage is minimal, testing effectiveness can suffer due to the sparse and potentially unreliable data produced by the FDM algorithm. In such cases, a detailed review of each flight might be required, as indicated previously.

How to share results & derive lessons learnt?

The final purpose of the testing exercise is to identify how the standard algorithm supports effective event identification, in a generalized way to different fleets and operating environments, independent of the operator. For this, it is important to focus on the evaluation of false positives and false negatives, against previously identified cases, as well as the evaluation of the basic metrics (rate of events) and their evolution, as well as their characteristics (what conditions trigger and in what volume, what severity levels...). To this end, in one way or another, the sharing of results and/or lessons learned is necessary as well as the comparison of results between operators. In addition, it may be of interest to maintain a framework for monitoring and sharing the results of standardised metrics, whether in the context of the same initiative or others.

Again, because of the sensitivity of the data in question, what can be shared depends on the governance framework and data and results sharing arrangements in the collaborative environment. Some considerations are given to this end:

What should be shared? Key considerations for blind-benchmarking & sharing results

The final purpose of sharing results is to encourage collaborative discussion around the standardisation of the proposed algorithm. Additionally, in a collaborative analytical environment, a data sharing scheme can favour the analysis of systemic risks, as well as derive lessons learnt or joint decision making, on common events or risk areas. For this purpose, as discussed in the testing subsection, it is interesting to share:

- **FDM event rate:** FDM events identified (number of detections) normalised per 1,000 or 10,000 flights (or flight hours) are used to provide an indicator of safety exposure that compensate changes in the volume of activity, thus being more robust for comparison and data sharing purposes.
- **FDM conditions triggered rate:** Beyond detected events, it might be of interest to provide rates of triggered conditions, in particular for the analysis of thresholds sensitivity for different operational scenarios and/or aircraft models

By means of bar charts or line graphs, the evolution of rates can be evaluated for the different (anonymised) operators that contribute data to the use case. All this in compliance with the de-identification rules agreed in the working group. Depending on the sharing environment (based on static reports or dynamic dashboards), rules for the development of such visualisations will be established to ensure understanding and protection of the results. For this, it will be necessary to consider not only the indicators provided, but also the contextual indicators provided. Some considerations in this regard:

- Visualisations focused on the FDM event rate can be accompanied by aggregation metrics, as well as the average FDM event rate, as long as the imposed de-identification rules are met. However, the generation of additional metrics, such as weighted average per traffic of the event rate, provides additional information that could favour the re-identification of the provided rates, by means of high availability information such as traffic volumes.
- The integration of FDM results with other data sources is another possible source of re-identification, which should be considered in the design of the results, as well as addressed in the proposed de-identification methods.

Beyond this, minimum representativeness should also be considered both to ensure reliable results and to avoid re-identification. Under limited number of flights (small sample of data), FDM event rates are an extremely sensitive indicator to small variations. As a result, a single event on a data sample of 1000 flights can have a big influence over the metrics, compared to a sample of 100.000 flights. This has an impact on the minimum sample of data necessary to obtain comparable results for each operator, at the different granularities necessary for the study (airport, temporal granularity - week / month...).

In addition, the comparison of rates between operators that do not guarantee a minimum of representativeness may also favour the re-identification of operators, especially in cases of low granularity. Example: Comparing event rates at airport level, with three operators contributing data (compliance with the rule of three) but with an unbalanced dataset. If one operator has its hub at the selected airport and the other two have few flights, due to the low volume at which their indicators are normalised, their rates could be close to 0 or extremely high. Without additional processing, it is easy to perform a re-identification exercise.

At this point, there are two fundamental points to be addressed: 1) Ensure data protection through de-identification techniques; and 2) Ensure a minimum sample of data that is representative of each operator, at the selected granularity. General considerations for the second topic are discussed in Annex B. High-level considerations for minimum representative sample in collaborative analytics of FDM data.

Who? Trusted third party as responsible for results aggregation & consolidation of lessons learnt

A trusted third party may be responsible for the aggregation of results by each contributing operator, whether or not it is responsible for the development of the analytics. The aim is to centralize in a single organisation the responsibility for safeguarding the de-identification of the results to be compared, in order to consolidate the lessons learnt for the approval of the testing exercise. All this with appropriate legal mechanisms to ensure proper management of the results and/or analysis of sensitive data. Such trusted third party could be one of the members under the umbrella of the governance framework and which the data owners can trust and which does not have a conflict of interest. Thus, an operator or authority that has a relationship with a data owner would probably not be acceptable. Instead, an example of trusted third party could potentially be an international organisation or an independent working group of experts under the right legal mechanisms established with each data owner. Some considerations for the two scenarios considered:

- **Scenario 1. Large-exchange data programmes.** In this case, this role is usually combined with the centralized analytics manager. The application of de-identification and data protection rules are applied before the generation of results, on the integrated flight data of all operators prior to the generation of metrics. This allows greater flexibility in the application of de-identification rules during analysis. Similarly, thanks to this governance framework, the organisation can drill-down to the necessary granularity and compare results between different operators, being able to differentiate differences in triggering conditions if any, as well as potential improvements to be introduced in the pseudocode in a more dynamic way.
- **Scenario 2. Aggregating results provided by individual operators.** In this case, the trusted third-party organisation is responsible for bringing together the results generated by each operator in its environments. In this case, it does not have access to the raw flight data, but to the results of a set of pre-agreed metrics, such as the rate of events, the number of flights considered, segregated by operational conditions (aircraft type, operating runway, if relevant...), the top conditions triggered and the safety levels, among others. The third-party organisation is responsible for consolidating the results and lessons learned from each company, as well as the main differences with their pseudo-codes of origin, and to propose a framework or collaborative analysis, on which the organisations can work without risk of re-identification of the operators. The de-identification rules, in this case, are applied a posteriori, on the results metrics generated by each operator. As noted above, this analytical environment is not desirable insofar as multiple variables are introduced that can generate noise in the comparison exercise, including the pre-processing rules of each operator's FDM technological environment.

These scenarios are neither exclusive nor unique, and more hybrid solutions for data management that enable collaboration can be defined. As an example, similar to what is promoted in large data sharing initiatives (e.g. Data4Safety), 'Demonstration datasets' could be generated with data from different airlines that has been de-identified and consolidated. A collaborative program could give the results of its metrics implemented on such demonstration datasets . Operators could have access to such flights and, potentially, run their own metrics on the demonstration datasets in order to evaluate the convergence of their methods with those of the collaborative programs and thus, creating an opportunity to derive lessons learnt for them and the programme towards standardisation.

The current environment in the EU appears to favour Scenario 1, with the opportunities presented by Data4Safety (D4S) providing an ideal foundation for further developing the standardization of events. Some first results have already been obtained during the Proof of Concept phase, demonstrating the value of the program in this regard.

However, consensus reached between ICAO, EASA, and the FAA, as outlined in AN-Conf/14, underscores the importance of going beyond regional initiatives, and creating a global framework for managing aviation safety data, to align and facilitate the coordination of regional voluntary programs. This global framework is needed to exchange results, best-practice and expertise to produce more accurate global safety information. As a result, while Scenario 1 provides a solid start, further efforts might be required to explore Scenario 2 for coordination of analytical efforts between regional exchange programmes, ensuring the international aviation community benefits from a unified, globally-coordinated safety data management system.

3. Conclusions

The European regulatory framework has favoured the progressive and effective implementation of FDM programmes as a fundamental part of air operators' SMS frameworks. Being a regulatory requirement for fleets with aircraft over 27,000 kg, there are many operators with long-standing, highly efficient, professionalised and highly tailored to the operational reality programmes. However, this is not true for all operators. Smaller operators find it complex to go beyond the standard implementation of the programme, which is usually provided by third-party companies. There are multiple reasons for this. Among them, the lack of available resources and investment. Despite the limitations of parameter availability in older aircrafts, data is progressively being enabled by new data frames. Technologically, there are more and more solutions that allow analytics on relatively large historical data (e.g., cloud services), in a cost-effective way. However, the organisational dimension is often a point of improvement, specifically in the capitalisation of analytical knowledge and its retention in companies.

Indeed, one of the main limitations to the effective progress of FDM programmes at the EU level is the capitalisation of the knowledge necessary for the definition, development, implementation and validation of algorithms adapted to the operational reality of each operator. Less mature operators opt for the use of standard algorithms provided by software vendors, with limited analytical capacity for their continuous evolution. Without deeper analytical knowledge on the operators' side and relevant guidance on what and how to implement FDM metrics, this additional step to adapt such algorithms to make them effective is challenging.

On the other side, more mature operators adapt the FDM algorithms as per their operation, based on the understanding of baseline pseudocodes provided by their providers and their own expertise. As a result of such adaptation, the definition of events and flight phases, the algorithms that represent these definitions, and even the parameters used, are not standardised, with each operator and software vendor using their own.

As a result, we have an industry with very disparate FDM programmes, with each operator adapting its FDM programme to its operational reality (size of fleet, volume of operations, available resources…), which can potentially hinder the advancement of the less mature ones. To address this, collaborative working groups such as EOFDM and EAFDM have developed guidance material to promote the implementation of FDM events that are relevant for risks common to the entire industry.

Furthermore, standardisation is a fundamental enabler in one of the industry's current analytical challenges, namely to favour data sharing between operators to facilitate common intelligence. Sharing results is complex without a common algorithmic framework through which benchmarking exercises are realistic.

All in all, this paper discusses potential approaches for the development of industry-agreed standard FDM event definitions. Specifically, it proposes to channel this both in the context of collaborative working groups, without joint analytical capabilities, as well as in the context of large data-exchange programmes, detailing the advantages and limitations of each approach. In addition, general recommendations for the development of the standardisation activity are provided. It is important to consider that the objective is the creation of an analytical standard framework that can on the one hand allow operators to critically and analytically adapt their FDM algorithms, and on the other hand favour the development of standard metrics for collaborative programmes with data sharing. To this end, a reference event definition card is proposed as a framework for documenting standard definitions. Additionally, considerations are provided on the analysis techniques and principles to be followed in standardisation activities, providing particularities for environments with and without data sharing.

Regarding next steps, this material developed under the DATAPP research project aims to provide a set of recommendations that could be considered and discussed in the framework of collaborative activities in the EU, such as the EOFDM Working Groups or EASA Data4safety programme.

4. Reference material

[1]. EASA & ALG, D-2.1 Development of the Case Studies, Available: https://www.easa.europa.eu/en/downloads/139336/en.

[2]. EASA, "Notice of Proposed Amendment 2024-02 - Regular update of the Air Operations rules", Available: https://www.easa.europa.eu/en/downloads/139589/en.

[3]. EASA, "Commission Regulation (EU) No 965/2012 Air operations", Available: https://www.easa.europa.eu/en/document-library/regulations/commission-regulation-eu-no-9652012.

[4]. EUROPEAN OPERATORS FLIGHT DATA MONITORING FORUM (EOFDM) WORKING GROUP C, "flight data monitoring analysis techniques and principles", Available: https://www.easa.europa.eu/en/downloads/134273/en

[5]. EUROPEAN OPERATORS FLIGHT DATA MONITORING FORUM (EOFDM) WORKING GROUP B, "Guidance for the implementation of FDM Precursors", Available: https://www.easa.europa.eu/en/downloads/119200/en

[6]. EUROPEAN AUTHORITIES FLIGHT DATA MONITORING FORUM (EAFDM), "Developing standardised FDM-based indicators", Available https://www.easa.europa.eu/sites/default/files/dfu/EAFDM__standardised_FDM-based_indicators_v2_Ed2017.pdf

5. Annex A. Illustrative example of FDM event definition card - Unstable Approach

The following table represents a non-exhaustive example of the content that should be presented in each of the standardised event definition cards, in this case focusing on the event "Unstable approach". Please, note that the table below do not represent any recommendation or guidance on the implementation of the proposed event and that it contains illustrative examples of conditions:

Table 5-1 Illustrative example of fields and conditions for the standardised event definition card

FDM Event:	Unstable Approach
Precursor metric name:	**Unstable Approach**
Metric definition	This indicator captures cases where the approach is still not stabilised or destabilised shortly before landing. The metric measures multiple criteria from FDM data.
Rationale	European aviation regulations require (for CAT operators) or expect all approaches to be flown as stabilised. An approach is classified as unstable if one or more of the predefined stabilisation conditions are not met. Unstable approaches are common contributors to accidents in the landing phase, such as runway excursions, tail strikes, bounced and hard landings. In AMC1 CAT.OP.MPA.115(a), EASA lists the criteria which should be satisfied in operations with aeroplanes. The output of the monitoring should allow the identification of the criteria that were not satisfied. It is also recommended to output the lowest altitude above airfield level (AAL) where this happened. It is also good practice to store the measurements of all the margins to the criteria for all the approaches regardless of whether the approach was stable or not (e.g., maximum speed during approach, maximum and minimum sink rate, maximum deviation from localiser / glideslope).
Metric scope / applicability	The KPI is typically computed for traffic flows, individual airports, or clusters of airports (selection/grouping based on size and/or geography).
Minimum set of flight parameters needed	Mandatory set of flight parameters: • Air ground • Pressure altitude Flight parameters that must be used if they are recorded[2]: • Vertical speed (Criterion 3) • N1 (Criterion 4) • … Additional exposure data: • Airport location (latitude, longitude) • Airport elevation

[2] Criteria without the required parameters will not be evaluated.

FDM Event:	Unstable Approach
Precursor metric name:	**Unstable Approach**
Data sources	• Flight Data Monitoring (FDM) programme • Airport infrastructure contextual datasets (navaids, RWY…)
Operational ranges for parameters	For the calculation of this metric, the following operational ranges should be applied in the data pipeline at approach window level: • N1 [0 %, 120 %] • Air ground [0, 1] • Pressure altitude [-1,000 ft, 70,000 ft] • …
Conditions triggering a detection	Any of the below criteria occurring below 1000ft HAT during 1st approach. • High Glideslope Deviation • Low Glideslope Deviation • Excessive Localizer Deviation • …

Algorithm or method / "trigger logic"	• The code filters those cases within the FDM dataset with the minimum set of parameters required to compute the metric:
	  ○ …
	• The touchdown timestamp and pressure altitude is determined. Two possible touchdown timestamps can be registered, with the earliest taken as the most accurate. The code computes all airport destinations per flight, classifying the airports as either landing or non-landing. Two sources for airports, the O/D parameter, and the computed destination (based on latitude/longitude). The airport selection logic is as follows:
	  ○ …
	• Within the approach phase window, it is checked when the aircraft crosses a set of height thresholds:
	  ○ IF current Height >= 500 ft and Height < 500 ft in the next 3 seconds;
	    ▪ 500 ft timestamp is recorded;
	  ○ IF current Height >= 50 ft and Height < 50 ft in the next 3 seconds;
	    ▪ 50 ft timestamp is recorded;
	• The Unstable Approach metric is composed of criteria (to be defined) that evaluate different flights parameters. The code assesses flights for each criterion, filtering out criteria without the required parameter or with parameters out of operational range (see "*Operational ranges for parameters*").
	• For each identified approach window, the analysis and evaluation of the criteria is triggered:
	  ○ Each criterion is defined within two different height ranges and two different severity thresholds. The number of criteria that meet the following conditions is checked and the last occurrence within the same height band per criterion is registered:
	    ▪ *Criterion 1:* High airspeed evaluated between the height ranges of (1,000-500] ft and (500-50] ft
	      • Airspeed greater than Vref + 20 knots for 3 seconds (1st threshold)
	      • Airspeed greater than Vref + 35 knots for 3 seconds (2nd threshold)
	    ▪ *Criterion 2:* Low airspeed between the height ranges of (1,000-500] ft and (500-50] ft

FDM Event:	Unstable Approach
Precursor metric name:	**Unstable Approach**
	• Airspeed lower than Vref - 5 knots for 3 seconds (1st threshold) • Airspeed lower than Vref - 10 knots for 3 seconds (2nd threshold) ▪ *Criterion 3:* Low vertical speed between the height ranges of (1,000-500] ft and (500-50] ft • Vertical speed lower than -1,200 fpm for 5 seconds (1st threshold) • Vertical speed lower than -1,500 fpm for 3 seconds (2nd threshold) ▪ …
Severity level assignation	• Severity level is assigned to each triggered criterion depending on both height range and severity threshold that is surpassed. The following is a non-exhaustive example of severity levels, from lower to higher severity: ○ *SL1*: Triggers of 1st threshold above or at 500 ft ○ *SL2A*: Triggers of 2nd threshold above or at 500 ft ○ *SL2B*: Triggers of 1st threshold below 500 ft ○ *SL3*: Triggers of 2nd threshold below 500 ft
Event identification	• An event is identified when: ○ Above or at 500 ft, at least 3 different criteria within the severity level 1 or severity level 2A are triggered, OR ○ Below 500 ft, at least 1 criterion within the severity level 2B or severity level 3 is triggered • Only one event is identified per approach and the event severity level will be the most severe of those registered
Additional considerations	• Flights must start and end on ground to avoid including corrupted flights to the analysis. • Flights must have a valid landing airport • …

6. Annex B. High-level considerations for minimum representative sample in collaborative analytics of FDM data

While FDM plays a very important role within the overall SMS of operators, FDM analysis are in many cases focused to the evaluation of sporadic and rare events, so the relevance of the results is highly dependent on maximising the availability of accurate and clean data, as well as the application of relevant statistical methods. By way of example, what minimum number of flights is relevant to assess, in terms of safety, to ensure that the conclusions derived regarding risk exposure are relevant? This is important in the analysis of the operator itself, and even more so, in collaborative analytics programmes to enable relevant comparisons.

As a result, one of the initial steps for FDM analytics is determining the minimum representative sample size necessary for accurate estimation of safety event rates and deriving meaningful insights from the data. The present chapter is based on methodologies and challenges faced in finding this minimum representative sample size which is essential for trustworthy and valid calculations of safety event rates.

Determining what constitutes an adequate sample size is not always easy. The process involves striking a balance between statistical rigor and practical considerations. Flight data captures large volumes of information in aviation sector but historical FDM event rates are often few, rare and sporadic. It can be difficult to estimate accurately their rates because most incidents such as mechanical failures, pilot errors or adverse weather impacts occur infrequently. Therefore, choosing a sample size that has both statistical significance as well as operational feasibility requires some understanding about statistical principles, previous trends within historical data sets and knowledge regarding FDM analytics specific contexts. Among them, it is important to take into account the following considerations in order to define a potential method for calculating the minimum reasonable sample:

- **FDM events are rare.** One of the primary challenges in FDM analytics is the rarity of safety events, in particular lagging indicators. Some FDM event rates often fall below 1 event every 1000 flights (e.g., 0.1%) and thus, traditional sampling methods may not be directly applicable. The fact that FDM events are so rare require for larger sample sizes to ensure that enough events are captured for reliable analysis, which can be complicated at certain granularities of analysis, due to low levels of exposure and difficulties in data collection.
- **Variability / Stability of FDM event rates:** Historical data provides an estimated FDM event rate, but these rates can vary over time due to changes in operational procedures, technological advancements, or external factors such as weather patterns. This variability complicates the sample size determination process, as the estimated rate may not always reflect current conditions accurately.
- **FDM events independency & homogeneity:** Can we assume that FDM events are independent and homogeneously distributed across samples? This is a critical assumption in sample size determination for particular statistical methods. In FDM analytics, it is crucial to ensure that safety events are independent and homogeneously distributed across different flights, within the scope of analysis.
- **We need to balance statistical accuracy & practical feasibility:** The larger the sample the better. Larger samples of data lead to more accurate estimation of event rates, but they also demand for more resources and/or might be impractical for some scopes of analysis. In that case, what it follows is a theoretical proposal for determination of the minimum sample of data required to lead to reliable

statistical measures but this criterion needs to be put into context by FDM analysts when performing their analysis, whenever it's not practical to apply it.

Potential methodological approach

There are multiple statistical methods for the estimation of the minimum sample necessary to derive reliable estimates, under a minimum confidence margin and minimum errors. Knowing the total population (number of flights in each analytical scope) and the historical proportion of each type of event, it is possible to calculate the minimum sample needed to assess how many future observations we need to calculate forward rates. Specifically, the following factors are needed to calculate the minimum sample size:

- What confidence level do we need? Typically, a 95% confidence level could be used.
- What is the expected proportion of safety events? It depends on the specific event. 1 every 1000 events or less.
- Population size: Total number of flights over a given period for a specific scope of analysis (geographic, airport…).

A potential approach would be to apply a normal distribution approximation, applying the sample size formula for proportions, as follows:

$$n = \frac{Z^2 \cdot p\,(1-p)}{E^2}$$

Where:

- n is the sample size.
- Z is the Z-value (the number of standard deviations from the mean) corresponding to the desired confidence level (1.96 for 95% confidence).
- p is the estimated proportion of safety events.
- E is the margin of error.

Example

Suppose you estimate the safety event rate (p) to be 0.1% (0.001), which is reasonable for very rare events. A 95% confidence level is required (Z=1.96) and a margin of error of 0.05% (0.0005).

Plugging in the values:

$$n = \frac{(1.96)^2 \cdot 0.001(1-0.001)}{(0.0005)^2} = 15{,}351$$

So, you would need a sample size of approximately 15,351 flights to accurately estimate the safety event rate with a 95% confidence level and a margin of error of 0.05%. Is it feasible for analytical purposes?

Large European airlines can operate >1,000 flights daily. A sample of 15,351 flights could be reached over some days, potentially allowing for weekly or 15-period analysis of reliable event rates at network level. Small European airlines, which might operate around 100-200 flights per day, make high-frequency analysis (e.g., weekly or monthly) impractical, from a statistical reliability standpoint. Margin of error or confidence levels should be reduced. Similarly, analysis at airport-level might be feasible for 10-15 days periods in major European hubs, while 3-month periods might be reliable from a statistical point of view for smaller European airports.

The choice of metrics for analysing and monitoring data, especially in aviation safety events, is highly influenced by understanding the underlying distribution. This understanding guides what is measured, how it is measured and how it is interpreted. For simplicity purposes, there are several cases when metrics are defined assuming a normal distribution, that is, statistical tests are designed under the assumption that data follows a normal (Gaussian) distribution. However, these metrics such as the mean or standard deviation will not provide reliable insights when the underlying data does not adhere to a normal distribution. This can be problematic, as different types of data distributions require different analytical approaches and metrics to capture the characteristics of the data accurately. In fact, for the purpose of analysing FDM events, normal approximation might not hold, as these are typically very rare events. If the normal approximation is questionable, an alternative approach to be considered is a Poisson distribution, which might be more suitable for modelling the number of rare events in large populations.

In essence, a Poisson distribution helps us understand and predict the probability of events occurring over a fixed period of time (or space), provided these events happen independently of each other and at a constant average rate. Let's dive into what this means and how it works with an example, and explore its applications in FDM safety analysis.

Imagine we're counting the number of times a specific safety event happens within a set period. This could be any safety precursor. If these events happen independently (meaning one event doesn't influence the occurrence of another), and at a constant rate, then the Poisson distribution can predict the probability of seeing a certain number of these events. Poisson distributions are typically used in the context of risk assessment methodologies and in the field of cybersecurity.

The Poisson distribution is built on some key assumptions, which are the following ones in the context of FDM analysis:

- **Individual events happen independently.** That is, the probability of one event doesn't affect the probability of another event. The occurrence of an event in one flight does not influence the occurrence of an event in another flight.
- **Infinitesimally small Probability (rare events):** The assumption that more than one event occurring in a very short interval is almost zero. This aligns well with certain types of safety incidents. For example, the likelihood of multiple distinct long landings occurring at the exact same moment is extremely low, allowing for the application of the Poisson model to predict the distribution of such events over time.
- **Constant average Rate:** In a given geographic / operational scope, the average rate of specific FDM safety incidents might remain relatively constant over time. This might be an arguable assumption, but it allows us to predict the expected number of incidents over a future period based on historical data.
- **Not limited Interval Length:** There's no practical upper limit to the number of safety incidents that could occur within a given time frame. This characteristic of the Poisson distribution makes it very well suited for modelling safety events, as it accommodates the unpredictable nature of operational risks, which can vary widely in frequency and severity**.**
- **Events Occur in a Continuum / large population:** FDM Safety events can occur at any point in time. Poisson distribution's flexibility in handling occurrences in a continuous time frame makes it particularly useful for this application.

To calculate the minimum sample size needed to estimate a rate using the Poisson distribution, the following factors need to be considered:

- **Estimated Event Rate (λ):** This is the expected rate of events per unit (e.g., per flight). It's the core of the Poisson distribution, helping to shape its curve.

- **Desired Confidence Level (Z-value):** This reflects how certain you want to be that the true rate lies within your confidence interval (e.g., 95% confidence level corresponds to Z = 1.96).
- **Margin of Error (E):** This is the maximum allowable difference between the estimated rate and the true rate.

The following formula can be applied for the calculation of the minimum sample under Poisson distribution:

$$n = \frac{Z^2 \cdot \lambda}{E^2}$$

Example

Suppose you estimate the safety event rate to be 0.1% (0.001), which is reasonable for very rare events. A 95% confidence level is required (Z=1.96) and a margin of error of 0.05% (0.0005).

Plugging in the values:

$$n = \frac{(1.96)^2 \cdot 0.001}{(0.0005)^2} = 15{,}366$$

So, you would need a sample size of approximately 15,366 flights to accurately estimate the safety event rate with a 95% confidence level and a margin of error of 0.05%.

In that regard, the "FDM analysis techniques and principles" document [4] also provides further details and guidance regarding statistics, which can additionally contribute to ensuring that the used data sample is representative and relevant for the analyses.

European Union Aviation Safety Agency

Konrad-Adenauer-Ufer 3
50668 Cologne
Germany
DATAPP (Digital Transformation - Case Studies for Aviation Safety Standards – Data Science Applications) | EASA (europa.eu)

Mail EASA.research@easa.europa.eu
Web www.easa.europa.eu

An Agency of the European Union