DOT/FAA/TC-25/33

Federal Aviation Administration
William J. Hughes Technical Center
Aviation Research Division
Atlantic City International Airport
New Jersey 08405

Aviation Safety Oversight Tools, Technologies, and Techniques

August 2025

Interim Technical report

U.S. Department of Transportation
Federal Aviation Administration

NOTICE

Technical Report Documentation Page

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
DOT/FAA/TC-25/33		

4. Title and Subtitle	Aviation Safety Oversight System Tools, Technologies, and Techniques
5. Report Date	August 2025
6. Performing Organization Code	ANG-E272
7. Author(s)	Tricia Gilbert, Jennifer Lamont, and Dr. Alan Bell
8. Performing Organization Report No.	
9. Performing Organization Name and Address	Systems Enginuity, Inc. (SEI) 8665 Sudley Rd #349 Manassas, VA 20110
10. Work Unit No. (TRAIS)	
11. Contract or Grant No.	692M15-23-F-00049
12. Sponsoring Agency Name and Address	U.S. Department of Transportation Federal Aviation Administration (FAA) Aviation Safety (AVS) Washington, DC 20591
13. Type of Report and Period Covered	InterimTechnical Report
14. Sponsoring Agency Code	AIR-020

15. Supplementary Notes

Research Sponsor – Michael Bartron, AIR-020
Research Project Manager – Dr. Huasheng Li, ANG-E272

16. Abstract

The FAA launched a Future of Oversight research initiative to meet regulatory and safety oversight challenges with the rapidly evolving U.S. National Airspace System with new aircraft technologies, emerging operations, and increasingly complex and global aviation industry operations and infrastructure. This initiative is intended to inform FAA policy and implementation of performance-based oversight, including regulations, systems, processes, and tools to ensure that FAA oversight aligns with public expectations for safety.

The Aviation Research Division (ANG-E2) at the FAA William J Hughes Technical Center was assigned two research tasks as an initial effort of the Future of Oversight research program, including Task 1 to *Conduct a Review of Oversight Methodologies* and Task 2 to *Conduct a Review of Oversight Systems* (i.e., tools, technologies, and techniques). An initial report, *Survey of Performance-Based and Risk-Based Oversight Strategies and Methods*, dated April 30, 2024, was delivered for Task 1.
This report for Task 2 presents a survey of existing and emerging safety oversight tools, techniques, and technologies from international aviation authorities, aviation regulatory organizations, and aviation industry. The survey also encompassed safety analytical tools and techniques as well as information management systems potentially beneficial for adoption in aviation safety oversight. Key observations, findings, and recommendations for further research are presented for oversight areas and tool capabilities for safety management system evaluation, safety risk analysis, safety information management, safety objectives and safety performance measurement, and oversight effectiveness evaluation among others.

17. Key Words	18. Distribution Statement
Aviation Safety -- Oversight -- Performance-based Aviation Safety -- Oversight -- Performance-based -- Tools, Technologies and Techniques	This document is available to the U.S. public through the National Technical Information Service (NTIS), Springfield, Virginia 22161. This document is also available from the Federal Aviation Administration William J. Hughes Technical Center at actlibrary.tc.faa.gov.

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	92	

Contents

Tables

Figures

Acronyms

Acronym	Definition
AAM	Advanced Air Mobility
ACAIS	Air Carrier Activity Information System
AFS	Flight Standards
AHCS	Aerospace Hazard Classification System
AI	Artificial Intelligence
AIDS	Accident & Incident Data System
AIR	Air Certification
ANSP	Air Navigation Service Provider
AOV	Air Traffic Safety Oversight
ASIAS	Aviation Safety Information Analysis and Sharing
ASRS	Aviation Safety Reporting System
ATM	Air Traffic Management
ATO	Air Traffic Organization
AVP	Office of Accident Investigation and Prevention
AVS	Aviation Safety
BPM	Business Process Modeling
BSEE	Bureau of Safety and Environmental Enforcement
CANSO	Civil Air Navigation Services Organization
CAST	Commercial Aviation Safety Team
CFR	Code of Federal Regulations
CICTT	CAST ICAO Common Taxonomy Team
DOE	Department of Energy
DOT	Department of Transportation
EASA	European Union Aviation Safety Agency
EI	Effective Implementation
FAA	Federal Aviation Administration
HIRMT	Hazard Identification, Risk Management, and Tracking
IATA	International Air Transport Association
ICAO	International Aviation Civil Organization
INC	Incident of Noncompliance
INCOSE	International Council on Systems Engineering
IOSA	IATA Operational Safety Audit Program
LLM	Large Learning Models
MBSE	Model-Based Systems Engineering
MKAD	Multiple Kernel-based Anomaly Detection
ML	Machine Learning

Acronym	Definition
MOR	Mandatory Occurrence Report
M-SCAIT	Maintenance Safety Culture Assessment and Improvement Toolkit
NAS	National Airspace System
NASA	National Aeronautics and Space Administration
NLP	Natural Language Processing
NRC	Nuclear Regulatory Commission
NTIS	National Technical Information Service
NTSB	National Transportation Safety Board
PDF	Portable Document Format
RBRT	Risk-Based Resource Targeting
RPA	Robotic Process Automation
SAS	Safety Assurance System
SM ICG	Safety Management International Collaboration Group
SMS	Safety Management System
SORA	Safety Oversight and Risk Assessment
SPI	Safety Performance Indicator
SysML	Systems Modeling Language
TLS	Target Level of Safety
TOKAI	Toolkit for ATM Occurrence Investigation
UK CAA	United Kingdom Civil Aviation Authority
UML	Universal Modeling Language
USOAP	Universal Safety Oversight Audit Programme

Executive summary

To address FAA's Future of Oversight research objectives, this report presents a survey of existing and emerging safety oversight tools, techniques, and technologies from international aviation authorities, regulatory organizations, and the aviation industry. The survey also encompassed safety analytical tools and techniques as well as information management systems potentially beneficial for adoption in aviation safety oversight. FAA Aviation Safety discussions and interviews regarding needs and challenges for future oversight systems and tools shaped the survey and literature review conducted for this report. The compilation of tools, including techniques and technologies, is organized by functional areas for safety oversight including safety management system evaluation, safety analytical tools and techniques, safety information management technology, safety objectives and performance measurement, and oversight effectiveness evaluation among other areas.

Key observations and findings based on the survey, including recommendations for further research, are highlighted below with a complete list in Section 10 of this report.

- Identification of high-level, goal-oriented safety objectives can be translated into both leading safety indicators and outcome-focused performance measures. Research to define safety performance indicators and measures holistically for FAA aviation safety oversight is needed along with a framework to allocate or trace higher-level objectives, indicators, and measures to lower level supporting oversight activities.
- The use of risk-based approaches as well as process modeling can inform the identification of safety objectives for oversight and connect those objectives to aviation safety risks and specific oversight activities and functions. In turn, this connectivity can promote safety oversight workforce understanding and adoption of performance-based oversight and how individual oversight work and activities contribute to overall aviation safety.
- Aviation safety oversight effectiveness and efficiency may benefit from establishing an oversight data exchange model to organize, standardize, and facilitate information sharing across regulatory and industry organizations. Such an exchange model could incorporate the FAA's Aerospace Hazard Classification System for management of aviation safety risk data.
- Artificial intelligence and machine learning (AI / ML) applications require data science expertise and organizational investment to curate and maintain data stores and guide iterative improvements to managed models. Oversight organization-specific needs, benefits, and readiness for AI / ML adoption should be examined with cross-cutting

needs captured in requirements that can be evaluated for centralized vs. distributed AI / ML infrastructure implementation.

- Artificial intelligence and machine learning techniques, coupled with big data amassed from aviation safety events and operations, offer insights into emerging and unknown areas of risk. Strategic planning is needed to ensure that agreements (or regulations) are in place with the industry to enable oversight data collection for new entrant technologies and operations.
- While precursor and root cause analysis can be conducted for individual aviation safety incidents and undesired safety events, there is a broader benefit to aviation safety when such data is examined collectively across regulated entities and operations for systemic patterns. Analytical tools, including AI / ML, should be explored for use in precursor and root cause analysis with actual oversight data to determine performance and benefits vs. investment in infrastructure and expertise as well as recommendations for additional oversight data collection.
- Research is needed to produce oversight guidance on how to analyze regulated entity manufacturing, quality, safety management, and change management processes for safety risks. As part of FAA oversight, an independent safety risk analysis of entity management processes can be conducted to ensure that processes are robust to errors of omission, commission, timing, and abnormal and non-nominal conditions such as production surges. Formal tracking and application of lessons learned from prior process lapses across all regulated entities can enhance safety risk analysis of new and modified management processes.
- Data on identified safety precursors and root causes should be combined with information on planned changes to aviation systems and operations. A predictive capability is needed to identify changes with greater risk to aviation safety to offer decision support in prioritizing such changes for proactive oversight.
- Approaches for independent audit of oversight effectiveness focus on the degree of implementation according to various maturity scales. To complement this approach, research into performance-based evaluation of oversight effectiveness is needed. Risk-based models, including barrier analysis, offer an opportunity to examine how oversight decisions and actions can be applied as risk mitigations. Further research is needed to determine if such models can provide a framework for not only evaluating individual oversight decisions and actions but also overall oversight effectiveness in reducing aviation safety risk.
- Value chain and other process modeling and analysis techniques could be explored to address the gap between the flying public’s expectations for aviation safety and current

safety oversight functions. Process modeling techniques offer an approach to capture the flying public's risk tolerances in comparison with the conduct and delivery of oversight services modeled in as-is and to-be versions of the oversight system.

- Further research is needed to investigate the potential identification of safety objectives and safety performance indicators for safety culture evaluation and to validate whether safety performance measurement can be used to predict and improve safety culture. This research should consider the evaluation of safety culture throughout all levels of an organization and the full supply chain where applicable.

Acknowledgments

The authors would like to acknowledge the valuable expertise and insights into FAA oversight needs and challenges provided by the FAA's Aviation Safety (AVS) Steering Committee for the Future of Oversight. Representatives across FAA Aviation Safety (AVS) organizations, including Flight Standards (AFS), Aircraft Certification (AIR), Office of Accident Investigation and Prevention (AVP), Air Traffic Safety Oversight (AOV), and Aerospace Medicine (AAM) contributed their expertise, ideas, and observations as part of AVS Steering Committee discussions and in various interviews held throughout this research project. Also recognized is the continued technical leadership and expertise of both the ANG-E2 technical monitor for this research and the AVS project sponsor, whose strategies and contributions shaped the direction of this study.

1 Introduction

The FAA launched a *Future of Oversight* research initiative to meet regulatory and safety oversight challenges with the rapidly evolving U.S. National Airspace System with new aircraft technologies, emerging operations such as advanced air mobility, and increasingly complex and global aviation industry operations and infrastructure. This initiative is intended to inform FAA policy and implementation of performance-based oversight, including regulations, systems, processes, and tools to ensure that FAA oversight aligns with public expectations for safety.

The Aviation Research Division (ANG-E2) at the FAA William J Hughes Technical Center was assigned two research tasks as an initial effort of the Future of Oversight research program:

> ***Task 1 – Conduct a Review of Oversight Methodologies***
>
> Survey performance-based safety oversight methodologies used by governments, industries, and organizations for safety oversight and surveillance. Conduct a literature review to understand to what extent regulatory authorities have conducted performance-based and risk-based oversight or surveillance. Collect and review safety oversight information to better understand the status of oversight in different domains worldwide including aviation, public transit, nuclear energy, petroleum, and food and drug industries, among others.
>
> ***Task 2 – Conduct a Review of Oversight Systems (i.e., Tools)***
>
> Survey safety oversight tools, technologies, and techniques used for the aviation domain, worldwide. Examine tools and systems for oversight data collection and analysis, risk prioritization, safety performance evaluation, and other functions, including example applications, automation, and recommendations for further research.

An initial report was delivered for Task 1: *Survey of Performance-Based and Risk-Based Oversight Strategies and Methods,* dated April 30, 2024. This report documents results of a literature survey of oversight mechanisms in aviation and other industries, focusing on performance-based and risk-based oversight approaches. Challenges for adapting FAA oversight approaches for performance-based methods and regulations were explored, including oversight of safety management systems in aviation.

For Task 2, this report focuses on a survey of aviation safety oversight tools, technologies, and techniques to inform Future of Oversight research goals and is intended as a companion to the report delivered for Task 1. As the FAA is aware of its internal tools and systems for oversight, briefly discussed in Section 1.1, the balance of this report examines existing and emerging aviation safety oversight tools beyond the FAA's environment.

1.1 FAA oversight tools

The FAA Aviation Safety (AVS) organization has an array of existing and evolving safety oversight tools and methods for compliance-based audits and inspections, risk-based prioritization of oversight, and targeted application of performance-based oversight of safety management systems. For example:

- *Safety Assurance System (SAS)*: This system is used within the FAA AFS organization to support oversight planning, data collection, and analysis. SAS facilitates work to assure that certificate holders are implementing safety risk controls and that these controls are effective. It is also used to identify changes in the environment and potential hazards or impacts not previously studied. The system is used to provide insight into oversight of SMS effectiveness, using component Data Collection Tools (DCTs) with questionnaires to assess if a certificate holder or applicant follows technical, procedural, and/or organization measures for applicable federal aviation regulations and standards.
- *Air Carrier Activity Information System (ACAIS)* / *Risk-Based Resource Targeting (RBRT)*: These tools support the collection and management of data by FAA aviation safety inspectors to support the assessment of risk. Criteria considered include the scope of the product and its associated risk; the stability of suppliers; the amount of work performed in-house by the approval holder; the approval holder's relationship with the FAA and compliance history; and the approval holder's company growth, among other items.
- *Hazard Identification, Risk Management, and Tracking (HIRMT) tool:* An agency-level tool intended to track certain aviation safety issues and associated resolution status. FAA Order 8040 provides criteria on what aviation safety issues must be tracked in HIRMT, such as existing and anticipated high-risk issues present in the NAS (FAA, 2023).
- *Aviation Safety Information Analysis and Sharing (ASIAS*): As both a program and a system, ASIAS supports the collection and analysis of a wide array of safety data from reported aviation incidents to aircraft flight records and aircraft sensor data. ASIAS capabilities include data search and visualization with metrics and trend analyses to examine potential systemic and emerging areas of risk.

- *Safety Oversight and Risk Assessment (SORA) tool*: Developed with AOV, SORA integrates and analyzes data on air traffic-related aviation safety events, air traffic operations, aeronautical information, and air traffic facility infrastructure and system performance. SORA provides trending and analytics for safety risk and performance indicators and cross-data search and visualizations.

A companion literature survey of FAA and other industry safety oversight methods describes these and other tools in more detail (Gilbert, Lamont, & Bell, 2024). As noted above, given the FAA's understanding of its existing oversight tools and techniques, the reset of this report examines aviation safety oversight tools, technologies, and techniques from international aviation authorities and aviation industry, among other techniques relevant to aviation safety oversight processes.

1.2 Purpose

The purpose of this report is to survey and present a compendium of tools, technologies, and techniques that can be used for aviation safety oversight functions, including oversight system effectiveness evaluation.

1.3 Definitions

The following definitions of performance-based and risk-based oversight are adapted from ICAO's Safety Management International Collaboration Group terminology (SM ICG), with additional direct sources as referenced below.

- **Assessment** – an appraisal of procedures or operations based largely on experience and professional judgment. (ICAO 9753, 1999)
- **Audit** – A systematic, independent, and documented process for obtaining evidence and evaluating it objectively to determine the extent to which requirements and audit criteria are fulfilled. (ISO 19011, 2018)
- **Aviation System** – The people, organizations, equipment, technology, and regulatory environment that interact to enable the development, production, operation, maintenance, and training associated with aircraft and aircraft components. (ICAO SM ICG, 2022)
- **Human Factors Principles** – Principles that apply to aeronautical design, certification, training, operations, and maintenance and which seek a safe interface between the human and other system components by proper consideration of human performance. (ICAO 9734, 2017)

- **Human Performance** – Human capabilities and limitations which have an impact on the safety, security, and efficiency of aeronautical operations. (ICAO Annex 6, 2018)
- **Safety Oversight or Surveillance** – A function performed by a regulator that ensures that an aviation organization complies with and uses safety-related standards, requirements, regulations, and associated procedures. This also includes an assessment of an organization's safety management. (ICAO SM ICG, 2022)
- **Performance Based Oversight** – Measurement of the performance of an organization's management system (e.g., SMS, QMS) in achieving the set safety objectives. It assesses the effectiveness of that management system. (ICAO SM ICG, 2022)
- **Safety Data** – a defined set of facts or set of safety values collected from various aviation-related sources, which is used to maintain or improve safety. (ICAO Annex 19, 2016)
- **Safety Management System** – A systematic approach to managing safety, including the necessary organizational structures, accountability, responsibilities, policies, and procedures. (ICAO Annex 19, 2016)
- **Safety Performance Indicator** – A data-based parameter used for monitoring and assessing safety performance. (ICAO Annex 19, 2016)

2 Approach

As aviation systems and operations become increasingly complex and regulatory approaches evolve, there is a growing need to adapt aviation safety oversight systems, tools, and techniques to provide rigorous safety risk management and safety assurance. The survey conducted for this study focused on existing and emerging tools, technologies, and techniques for aviation safety oversight.

2.1 AVS discussions

To guide the vision and goals for the Future of Oversight research initiative, the FAA formed a steering committee comprised of Aviation Safety (AVS) oversight experts. AVS Steering Committee discussions in 2024 regarding needs and challenges for future oversight systems and tools shaped the survey conducted for this report along with findings from AVS interviews previously conducted for the companion report, *Survey of Performance-Based and Risk-Based Oversight Strategies and Methods*. The following recaps findings on FAA oversight needs and future challenges based on those previous interviews with updated observations from the AVS Steering Committee discussions:

1. Regulatory framework changes necessitate changes in oversight methods and tools

(a) Oversight resources, methods, and tools need to accommodate oversight of both prescriptive and performance-based regulations, as regulatory changes expand to incorporate more performance-based elements such as a Safety Management System (SMS) while retaining prescriptive regulations for safety-critical operations and systems (multiple AVS organizations).

(b) Risk-based oversight strategies should include both compliance-based and performance-based elements, where compliance issues and safety performance, including SMS effectiveness issues, inform risk-based priorities for oversight focus areas, frequencies, and regulated entities (AFS, AVS Steering Committee).

(c) Specific approaches to defining safety objectives, measuring safety performance, and evaluating safety management effectiveness must be defined and validated along with strategies to assure industry self-correction is occurring and effective (AIR, AVS Steering Committee). SMS effectiveness evaluation should account for safety culture and how well the management system as implemented meets safety objectives (AVS Steering Committee).

(d) Oversight methods and approaches should be agile and mature over time – for example, initially focusing on assessing the implementation and application of specific processes such as SMS and later focusing on measures that reflect the effectiveness of safety risk management processes (AIR, AVS Steering Committee).

(e) Oversight of SMS effectiveness will require a "look-back" approach similar to root cause analyses, where the approach to issue identification and corrective action are supplemented by an approach to determine which SMS or other management system gaps and process lapses allowed the failure to occur (AIR).

2. Oversight strategies must incorporate surveillance on changing operations, processes, and technology and associated safety and quality management

(a) The aviation industry and technologies are changing quickly, and as new and modified technologies and operations are introduced in the National Airspace System (NAS), the FAA needs to develop a comprehensive and independent understanding of those new technologies and operations, and especially safety-significant changes, to provide effective oversight (multiple AVS organizations).

(b) Oversight needs to incorporate predictive tools, leveraging data from surveillance, to anticipate potential safety issues and where oversight should be applied (AVS Steering Committee).

(c) Assumptions and risk mitigation identified during the certificate holder or production approval process need to be re-evaluated for validity after initial approval, especially with new and changed technology and systems. This drives a need to collect, track, and monitor more extensive and more frequent data for continuous safety assurance (AVS Steering Committee).

(d) An oversight capability is needed to track risk assessment and risk mitigation data produced by regulated entities, and to validate that the safety risk assessed and need for mitigation is consistent with independent FAA risk modeling and analysis and a process to adjudicate differences. A common hazard taxonomy is fundamental to implementing this concept as well as effective and efficient safety risk management (AVS Steering Committee).

(e) Oversight activities should analyze regulated entity manufacturing, quality assurance, and other management system processes and procedures to evaluate safety and verify outputs are valid and useful for continued safety assurance (AVS Steering Committee).

3. Cross-organization oversight integration and industry collaboration is needed for effective and efficient risk-based oversight

(a) Oversight methods and tools that can be tailored for specific oversight needs within AVS organizations can also adapt to address provisions for performance-based oversight as related oversight processes are expanded and matured. As oversight tools become more integrated to meet needs across AVS, there is a greater need for data sharing among oversight organizations to realize the benefits of those integrated tools in risk-based oversight planning and prioritization (AIR).

(b) Oversight should take a holistic approach that accounts for how activities of different AVS organizations impact one another and the same regulated entities engaged in multiple oversight activities (AIR).

(c) Some regulated entities lack knowledge and understanding of SMS, certification, and FAA continued operational safety policy and processes; the FAA will need to improve and expand resources for industry education and collaboration, as this area currently consumes many oversight resources (multiple AVS organizations).

(d) Oversight tools and methods are needed to assess the impact and value of collective oversight processes such as surveillance, with traceability to rules and regulations so that potential improvements can be evaluated (AVS Steering Committee).

4. Workforce development and both FAA and industry safety culture investments are needed to facilitate successful performance-based and risk-based oversight implementation

(a) An evolution of the oversight workforce is needed, focusing on expanding knowledge, experience, and confidence in applying performance-based oversight (multiple AVS organizations).

(b) Expectations for the required workforce development timeframe to adopt performance-based oversight need to be realistic in planning for expected outcomes and benefits (AIR).

(c) Workforce training needs to address changes in approach and mindsets between compliance-focused and performance-based oversight (AFS), including the use of systems-level thinking to evaluate the effectiveness of the regulated entity management systems and to interpret and respond to collected oversight data (multiple AVS organizations).

(d) Actions focused on achieving desired transparency and information sharing are needed both within the FAA and between FAA and regulated entities; this could include inputs to specific oversight methods as well as other activities to drive a positive safety culture (AVP, AIR). Success cases and lessons learned can also be documented and shared as part of this activity (AFS). Oversight performance results, and internal assessments, could be made transparent, with public information sharing to build trust (AVS Steering Committee).

(e) When the aviation system appears to be safe, it can be difficult to motivate the industry to invest in continued safety improvements and to fully promote SMS and safety culture principles within its operations and processes (multiple AVS organizations)

2.2 Literature review

Findings from Future of Oversight steering committee discussions and AVS interviews were used to inform a literature survey of aviation safety oversight tools, technologies, and techniques (collectively, "*tools*" for ease of reference). A literature review was initiated in June 2024 which resulted in a compilation of over 100 publications, online resources, and other media addressing aviation safety oversight tools and applications (see Section 11). The literature review focused primarily on material published in the past five years, with an effort to identify material addressing best practices, lessons learned, and applicability within aviation safety. Tools from international aviation authorities, aviation safety organizations and associations, and industry were researched for applicability to safety oversight. Certain tools are directly produced or used by aviation safety oversight or regulatory organizations, whereas other tools from the aviation industry and industry associations support internal safety oversight practices. The literature review also encompassed safety analytical tools and techniques as well as information management systems that are not aviation industry-specific but have existing or potential applications for aviation safety oversight. An organizing principle was applied within this report to group tools according to their primary functions as follows:

- Safety management system evaluation
- Safety analytical tools and techniques
- Safety information management
- Safety objectives and performance measurement
- Oversight effectiveness evaluation
- Process modeling and automation
- Safety culture and human performance assessment

2.3 Safety management pillars

FAA Order 8000.367C establishes requirements for the AVS Safety Management System, consistent with agency-wide SMS policy in FAA Order 8000.369C. Though 8000.367C is undergoing a revision, it is anticipated that AVS needs for aviation safety oversight tools, technologies, and techniques will continue to align with the four pillars of SMS:

- *Safety Policy:* defines the need to establish requirements, methods, and commitment to achieve a defined set of safety outcomes.
- *Safety Risk Management:* Defines the action to analyze performance, identify and evaluate hazards and associated safety risk, and take action to control and monitor hazards. This section references a more detailed Safety Risk Management policy in FAA Order 8040.4.
- *Safety Assurance:* processes oversight actions to verify that intended safety performance is met; actions defined include data analysis; employee/stakeholder reporting systems; monitoring/investigation, evaluation, and auditing; corrective action identification and tracking; and management review.
- *Safety Promotion:* establishes mechanisms for communicating safety information; ensures the workforce has the requisite competencies and training to perform their mission; and reinforces a positive safety culture.

2.4 Mapping

The table below lists the tools, technologies, and techniques surveyed in this report for existing and potential use in aviation safety oversight. Tools are mapped to the numbered findings from

AVS interviews and AVS Steering Committee discussions in Section 2.1 and to the four SMS pillars outlined in Section 2.3. As noted in Section 2.2, an additional organizing principle was applied within this report to group tools according to their primary functions.

Table 1: Mapping SMS pillars to surveyed tools

Tool / Technology / Technique	**AVS Discussions: Numbered Findings from 2.1**	**SMS Pillars from 2.3**			
		Safety Policy	**Safety Risk Mgmt.**	**Safety Assurance**	**Safety Promotion**
SMS EVALUATION					
SMS Evaluation	1(a), 1(b), 1(c), 1(d), 1(e), 3(b)			x	
Aviation Industry SMS Tools	1(a), 1(b), 1(d)	x	x	x	x
SAFETY ANALYTICAL TOOLS					
Artificial Intelligence / Machine Learning	2(b), 2(d)		x	x	
Precursor and Root Cause Analysis	1(c), 1(e)		x	x	
Barrier Modeling	1(e), 2(d), 3(d)		x	x	
Occurrence Analysis	1(c), 2(a)		x		
Sequential / Simultaneous Hazard Modeling	2(d), 3(b)		x		
INFORMATION MANAGEMENT					
Big Data	2(b), 2(c), 3(a)		x	x	
Blockchain	2(c), 3(a)			x	
Safety Taxonomies	2(c), 2(d), 4(d)		x	x	
Data Exchange Models	2(c), 3(a)		x		x
Data Visualizations and Dashboards	2(b), 3(a)		x	x	x
SAFETY OBJECTIVES & PERFORMANCE					
Safety Objectives	1(c), 1(b)	x			
Safety Performance Measurement	1(c), 1(b)		x	x	
OVERSIGHT EFFECTIVENESS EVALUATION (see also Safety Performance Measurement)					
Independent Audits	1(b),1(c), 3(d), 4(d)			x	
Stakeholder Surveys	1(b), 1(c), 3(d), 4(d)			x	
Benchmarking	1(b), 1(c), 3(d), 4(d)	x		x	
PROCESS MODELING & AUTOMATION					
Model-Based Systems Engineering	3(b)		x	x	
Value Chain Analysis	3(d)		x		x
Business Process Modeling	2(c), 2(e), 3(b), 3(d)		x	x	x
Robotic Process Automation	2(c)		x	x	
SAFETY CULTURE & HUMAN PERFORMANCE					
Safety Culture Assessment	1(c), 4(a), 4(d), 4(e)	x	x	x	x
Human Performance and Interaction Assessment	4(a), 4(b), 4(c)		x		

3 SMS evaluation

Several checklist-style tools have been introduced with the goal of evaluating the effectiveness of a Safety Management System in aviation. There are several common threads that run through the documents, all of which were produced between 2019 and 2023. One of the more important threads is a collection of statements describing the intent of oversight, with two overarching goals: auditing for compliance and evaluation of effectiveness.

3.1 ICAO SMS evaluation tool

Several international aviation safety authorities have developed SMS evaluation tools based on version 2 of ICAO's SMS evaluation tool introduced in 2019 by the Safety Management International Collaboration Group. According to SM ICG guidance on SMS evaluation, compliance, and performance considerations factor into the evaluation of SMS effectiveness. Attributes of safety risk management, assurance, policy, promotion as well as internal and external interface management are presented as indicators (or statements) of compliance and performance. Guidance is provided on assessing the maturity of an organization's SMS as present, suitable, operating, or effective along with examples of "what to look for" when interviewing a service provider or examining service provider artifacts. Effective implementation of safety risk management entails, for example, "continuously and proactively identifying hazards related to [the service provider's] activities" and "safety investigations identify causal/contributing factors that are acted upon." Effective safety assurance exhibits safety performance indicators (SPIs) that "are demonstrating the safety performance of the organization and the effectiveness of risk controls based on reliable data" where "SPIs are reviewed and regularly updated to ensure they remain relevant."

In March 2024, ICAO reported user feedback on its SMS evaluation tool as part of a broader survey of SM ICG products (ICAO, 2024). The survey found that the SMS evaluation tool was used for familiarization, training, and guidance, and in practical applications to assess the level of effectiveness of SMS implementation within an organization. Outside of ICAO's survey, FAA representatives within AVS observed that attempts to directly apply the SMS evaluation tool for airworthiness organization SMS implementation proved challenging and that additional guidance and tailoring were needed for specific use cases.

3.2 EASA Management Assessment Tool

The European Union's Aviation Safety Agency's (EASA) Management Assessment Tool, Edition 2, is intended to support an evolution of oversight from traditional compliance-based oversight to performance-based oversight. It is also intended to be used as guidance for oversight of both the initial implementation of an SMS as well as continuous oversight once an SMS is operational. It can also be used by external oversight authorities or internal safety professionals in preparation for audits and inspections by the oversight organization (EASA, 2023). The assessment tool offers guidance on 27 items that should be inspected during an audit. Each item has four levels, with a high-level description of each in the table below.

Table 2: EASA Management System Assessment Tool Maturity Levels

Level	Description
Present	There is evidence that the relevant item is documented within the organization's Management System Documentation
Suitable	The relevant item is suitable based on the size, nature, complexity of the organization, and the inherent risk in the activity
Operating	There is evidence that the relevant item is in use and an output is being produced
Effective	There is evidence that the relevant item is achieving the desired outcome and has a positive safety impact

Each of the 27 items are based on requirements and are phrased as "shall" statements. For instance, "The safety policy shall reflect organizational commitment regarding safety, including the promotion of a positive safety culture." The inspector then evaluates the maturity level by assessing whether the specific elements associated with each maturity level definition are met or not. In addition to the maturity level definitions, there is also a guidance section entitled "What to look for" to add details and examples of things an inspector could use to support an affirmative or negative assessment of each criterion.

EASA documentation includes guidance that states the tool is not intended to be used as a checklist, implying that it can be used as part of an evaluation process but should not by itself be considered a full evaluation of SMS implementation and effectiveness. Of note is that the requirements in this tool do not explicitly call for establishing safety objectives that are tied to measurable performance data such as the number of accidents or incidents, risk exposure, or monetary loss. This could point to a gap in the full evaluation of the effectiveness of an organization's SMS in terms of its influence on the ultimate objectives of maintaining or reducing the amount of loss or harm experienced in conjunction with operations.

3.3 UK CAA SMS Evaluation Tool

The United Kingdom Civil Aviation Authority (UK CAA) SMS Evaluation Tool v7 is intended to evaluate the effectiveness of an SMS. The tool is a checklist based on the SM ICG template and others in use, including the EASA checklist. The tool uses the same maturity levels as its SM ICG/EASA counterparts, with elements described as "Present, Suitable, Operating, and Effective." For each element of the assessment, the evaluator checks the box for any of these terms, referred to as “markers.” An interesting variation in the application of these terms is that the UK version of this checklist introduces “Initiating” as another maturity level and then combines “present and suitable” into a single definition. The definitions also distinguish between the other two terms using a common metric: *effectiveness*. That is, if an element is present and suitable, the distinction between the remaining two terms depends on the evaluator’s assessment of effectiveness. If the evaluator subjectively considers the element *effective*, then it is understandably rated as “effective.” However, if the evaluator considers the element *ineffective*, the evaluator, the element must be rated as “operating.”

The UK CAA checklist includes the same elements as its predecessors, and also includes guidance for “what to look for,” providing evaluators with additional information to help them decide which of the markers apply in each case. There is also a text box available for explanations regarding how each element achieved the level it is assessed at.

Another interesting variation introduced by the UK is a summary section at the end of the checklist. There is a section labeled “SMS Evaluation Risk Picture” to summarize the “Top 5 safety risks or issues.” While the UK SMS concept is based on ICAO’s approach which in turn bases risk management on hazard analysis, the summary section does not ask for the Top 5 “hazards,” but instead, “Safety Risks or Issues.” Presumably, the intent is to list the five hazards that are assessed as being the highest risk. There is also an overall SMS Evaluation Summary for each of the pillars of SMS, plus human factors. Finally, there is an overall assessment, with each summary using the same four markers as the individual element evaluations.

3.4 MITRE guidance on safety management effectiveness

MITRE’s May 2021 report, Evaluating Safety Management Effectiveness, offers a summary of Safety Management Systems with emphasis on aviation safety (Hollinger, 2021). The document describes the four-pillar framework of policy, risk management through hazard analysis, assurance, and promotion as being a best practice. It then offers the best discussion found among the documents reviewed herein of the distinction between auditing for compliance and evaluating effectiveness. MITRE describes auditing as a “necessary and vital step in determining an organization’s compliance,” but continues to say that “auditing falls short in determining whether or not the program is effective,” and that “a multi-layered approach, using both audits and evaluations, is necessary.” The remainder of the paper focuses on evaluating effectiveness.

The MITRE paper also describes the benefits of implementing a safety management system, and these benefits are ultimately the justification for investing the resources necessary to implement, maintain, and operate a safety management system. Among the benefits listed are the key *safety objectives* of any SMS: reduced injuries to people and reduced damage to property.

The paper recommends auditing to ask yes or no questions for each pillar of the SMS. As one example, the paper proposes that evaluating safety policy might involve answering a question like, “Is the policy widely understood and trusted by all levels of management and frontline employees.” If the answer to this and other related questions is yes, then this is an indication that the SMS policy is *effective*. The paper goes on to propose multiple questions of a similar nature for each of the four SMS pillars and establishes a theme that is common to virtually all of the reviewed safety oversight evaluation activities: SMS evaluation is a qualitative process. For example, the paper outlines safety panel membership composition and proper assignment of risk classifications as considerations for effective safety risk management. For safety assurance, effectiveness in that process is gauged by questions such as whether mechanisms are in place for ensuring the effectiveness of risk controls (including re-examination of ineffective controls), employee reporting of safety concerns, and learning from failures.

3.5 SMS gap assessment

Aviation regulators such as the FAA can play a crucial role in facilitating SMS implementation within organizations that are required or choose to implement these systems. One means of support is providing SMS gap analysis assessments, evaluating how well an organization's current safety management practices align with the requirements and best practices of SMS standards with the goal of identifying deficiencies or gaps in the safety management system and allowing organizations to develop strategies to address these shortcomings. An assumption for these activities is the existence of clear standards and guidelines for SMS, addressing requirements for safety policy, risk management, safety assurance, and safety promotion.

Gap analysis involves comparing an organization’s current safety management practices against the established standards and requirements. Aviation regulators can offer this service directly or guide organizations on how to conduct their own assessments. The process typically includes:

- Documentation Review: Evaluating existing safety policies, procedures, and records to identify discrepancies with SMS requirements.
- Interviews and Observations: Engaging with personnel to understand how safety practices are implemented in reality compared to the documented procedures.
- Risk Assessment: Identifying areas where risk management practices might be lacking or insufficient.

After conducting or facilitating an SMS gap analysis, aviation regulators can offer detailed feedback and actionable recommendations that may help organizations address specific deficiencies and improve their SMS implementation. Offering to perform or facilitate an SMS gap analysis can support the development of a collaborative environment to build trust and promote open and honest communications.

Some Air Navigation Service Providers (ANSPs) have developed tools to facilitate the conduct of SMS gap analyses. For example, the UK CAA has developed tools to assess SMS gaps, with tools tailored to complex vs. non-complex organizations, according to the categories outlined in Table 3 (UK CAA).

Table 3: Example SMS gap analysis tool topics

Category	**Subcategory**	**Evaluation Elements**
General	SMS scope and implementation	Assessing the scope of the organization and its activities including contractors providing safety services to the organization; evaluating SMS in the context of size, nature, and complexity of the organization and its implementation plan
Safety Policy	Management Commitment	Assessing safety policy alignment with requirements and signed / approved by an Accountable Manager. Ensuring required resources are allocated and policy is communicated, including actions taken when the policy is not followed.
Safety Policy	Safety Accountability and Responsibility	Assessing policy to identify accountable functions and persons, including documentation and communication of all safety responsibilities, accountabilities, and authorities
Safety Policy	Appointment of key safety personnel	Ensuring policy identifies a safety manager to be the responsible person for the implementation and management of an effective SMS.
Safety Policy	Coordination of Emergency Response Planning	Ensuring that an emergency response plan addresses orderly and efficient transition to emergency operations and return to service.
Safety Policy	SMS Documentation	Ensuring SMS documentation is developed and maintained.
Safety Risk Management	Hazard Identification	Evaluating the formal process to identify aviation hazards (including reactive, proactive, and predictive data collection).
Safety Risk Management	Safety risk assessment and mitigation process	Evaluating formal processes to ensure analysis, assessment, and control of safety risk in operations to as low as reasonably practical.
Safety Assurance	Safety performance monitoring and measurement	Assessing organizations approach to develop and maintain the means to verify the safety performance of the organization and validate the effectiveness of risk controls.
Safety Assurance	Management of Change	Evaluating the organization’s approach to develop and maintain a formal process to identify changes; describing arrangements to ensure safety performance before the change is implemented; and eliminating or modifying safety risk for controls no longer needed, effective, or available.
Safety Assurance	Continuous improvement of SMS	Assessing the organization's formal process to identify and improve substandard performance of the SMS

Category	Subcategory	Evaluation Elements
Safety Promotion	Training and Education	Evaluating the organization’s approach to developing and maintain a safety training program, ensuring persons are trained and competent to perform the SMS duties
Safety Promotion	Safety Communication	Assessing the organization's formal means of developing and maintaining safety communications.

Although the tool summarized above was developed for the evaluation of regulated organizations’ SMS gaps, a similar structure and topics can be tailored and applied for internal evaluation of an oversight organization’s performance to its SMS, identifying gaps and opportunities for improvement.

3.6 Aviation industry SMS tools

3.6.1 Spirit Airlines

A case study on Spirit Airlines' adoption of MA Software Systems SMS Suite provides a backdrop for describing and demonstrating the capabilities of the SMS software designed to show compliance with Federal Aviation Regulations. It is intended to facilitate compliance with the four pillars of an SMS by providing an interface for users to report incidents. The software then uses a variety of algorithms to extract hazards from text-based narratives and offers options for managing risk associated with those hazards. The risk assessments are qualitative in nature and rely on subjective interpretations of categorical severity and likelihood scales. The software includes audit functionality that provides a series of questions and checklists for generating findings and tracking corrective actions by operational divisions. The presentation also describes an Artificial Intelligence capability that is intended to rely on language models to identify hazards and provide risk insights to subscribers (MA Software Solutions, 2024).

3.6.2 Delta Airlines

According to its SMS website, Delta Airlines initiates safety risk management as part of its change management process. In addition, Delta introduced *FlightPulse*, “a real-time performance review tool that lets pilots see hazard-related risks and make adjustments mid-flight” (Delta Airlines, 2024). Delta’s SMS also includes safety assurance processes, with audits of compliance and effectiveness through an internal evaluation program. Process audits play a role in maintaining the health of operationally critical policies and procedures. Delta also conducts regular quality assurance audits at domestic and international locations to evaluate ground and maintenance operations. Audit results are incorporated into the SMS and reviewed by leaders every month. Delta also receives independent, third-party safety audits through the FAA, IATA's Operational Safety Audit program, and the U.S. Department of Defense.

3.6.3 United Airlines

United Airlines has developed a mobile application that allows every employee to report safety incidents and concerns from a work or personal device. According to its website, nearly 50% of all safety incident reports are submitted through this safety application (United Airlines, 2024). Additionally, the airline has five operational-specific safety reporting programs, all of which are voluntary and non-punitive. United relies heavily on a just culture and voluntary reporting to identify hazards that could cause accidents. However, there is little detail within United's public SMS information as to how management uses the voluntary reports, or how the identified hazards are addressed through United's safety risk management process.

3.7 Sampling changes under SMS

Some regulatory authorities, such as the UK CAA, have implemented a sampling approach to assess confidence in an organization's processes that define and manage risk for a specific type of proposed change. The UK CAA evaluation entails six steps (UK CAA Safety and Airspace Reguation Group, 2018):

- Step 1: Review of a change's nature, scope, and impact
- Step 2: Identification and assessment of hazards and consequences
- Step 3: Risk evaluation and acceptance
- Step 4: Review of identified risk mitigation measures
- Step 5: Review justification and supporting evidence
- Step 6: Assessment of the assurance plan for managing residual risk

Change sampling may be used as part of a larger oversight strategy that determines the level of aviation authority involvement in evaluating the outputs of a regulated entity's change management system and SMS. Higher levels of involvement, including requests or requirements for oversight organization pre-approval of change packages, correspond to the regulator's confidence in the organization performing the change to successfully manage that change including potential safety impacts. This concept, adapted from UK CAA change oversight guidance, is provided here as Figure 1.

A factor in both assessing an organization's change management system and in determining regulatory involvement in change management is sufficient technical understanding of the organization and associated system or service under change. This includes an understanding of safety significance of a change both by the organization making the change and the applicable oversight organizations.

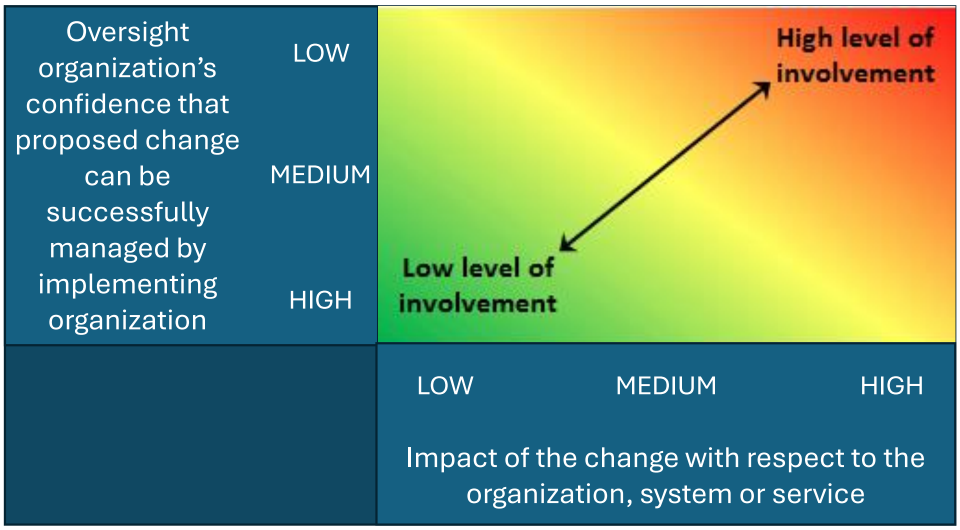


Figure 1: Regulatory oversight involvement in change safety evaluation

4 Safety analytical tools and techniques

4.1 Artificial Intelligence (AI) / Machine Learning (ML)

Artificial Intelligence and Machine Learning (AI / ML) employ data science techniques to analyze complex data sets to uncover patterns and problem solutions. Machine learning excels at analyzing large, intricate, and diverse datasets in ways that are challenging or impossible for the human mind, such as through n-dimensional analysis. Within aviation safety oversight, there are growing and complex problems that may benefit from AI / ML advances to analyze data, including:

- *Classification*: Organization and classification of incidents to support analysis (e.g., changes over time); this can be used by an oversight organization to independently assess the overall performance of a safety management system through evaluation of leading safety indicators and safety outcomes.
- *Cluster / Data Pattern / Factor Analysis*: Extracting data patterns that may be representative of aviation incident factors of interest such as contributing or mitigating

factors, emerging factors, or outcomes; this can be used by an oversight organization to identify and prioritize factors for oversight focus in a risk-based manner.

- *Outlier Detection*: Leveraging defensible and unbiased data points to identify off-nominal events and support proactive safety oversight decision-making (National Business Aviation Association (NBAA), 2021); this can be used by an oversight organization to proactively identify changes that may reflect new and emerging risk.
- *Risk Assessment*: Support risk assessments by accessing safety data sets and simultaneously evaluating the influence of multiple variables to identify relationships and dependencies between hazards and occurrences at all levels of severity. ML can be used to augment current data sets by analyzing known information within incident and accident reports, including aircraft type, number of engines, type of operation, phase of flight, number of seats, and pilot experience, among others, to estimate missing data fields in other reports. Artificial Intelligence can then be used to generate multiple alternatives for risk mitigation along with probabilistic estimates of their influence on risk, either individually or in combination with other controls. When used in an oversight capacity, ML and AI can evaluate the results of a risk assessment for concurrence, or to validate historic assessments using recent data. AI may also be used to identify anomalous events that could indicate a need for oversight action or investigation to proactively manage emerging risks before they culminate in accidents.

As NAS operations are projected to exponentially grow, the need to automate and move towards a more proactive and predictive SMS becomes increasingly necessary as discussed in a concept for an *"In-time Aviation Safety Management System"* (Ellis, et al., 2022). Aviation operations growth magnifies operational complexity given increasingly dynamic interactions between humans and technology with unclear impacts. To both understand and mitigate safety events, more domain expert knowledge of the underlying system is required and is quickly becoming unreasonably challenging without the aid of advanced data analysis techniques. AI / ML can play a role in augmenting and automating data analysis to support safety oversight, accommodating additional safety data volume and complexity. The subsections below address certain prerequisites for using AI / ML followed by opportunities to support safety oversight via classification, cluster analysis, outlier detection, and risk assessment techniques.

4.1.1 Prerequisites for data analyses

A fundamental prerequisite for establishing an AI / ML capability entails understanding and preparation of data sources chosen for analyses. Aviation safety and operational data must be properly understood, cleaned, and consolidated in an efficient and accurate manner for ingestion by analytical tools. Without a solid data foundation, organizations are left with inaccurate, unusable models. It is vital to first spend resources mindfully planning, organizing, and automating data source aggregation before introducing data analysis. Using machine learning tools requires a commitment to constant development and maintenance of not only ML tools but also data ingestion automation to ensure reliable data sources are available for data analysis.

Once data collection and aggregation methods are in place, there must also be a commitment to continuous auditing and maintenance of the data cleaning and loading process. Although much of data engineering may be automated (e.g., refer to robotic process automation in 8.4), it is common for data to be loaded incorrectly as errors are introduced into the data source or technologies evolve. These mistakes can be easily overlooked without quality checks from analysts. Continuous comparisons between a cleaned data warehouse versus the raw source of the data should be implemented to identify any errors. Only once an organization has a reliable data warehouse in place can they begin to consider data analytics.

It should be noted that machine learning models are only as productive and accurate as the data provided. For supervised models, it is essential that the model's training data set accurately represents real-world data that the model will analyze. Trade-offs regarding the specificity versus the size of the training set must be carefully weighed. Generally, the more data used for training, the better the predictive power the model will have given a higher likelihood of sufficiently representing the data. In reality, large quantities of data are often unavailable, and if a training set is too small, there is a risk that the model will not understand the full scope of the data. These constraints must be thoughtfully weighed to support the desired use case for analysis.

As an additional perquisite for establishing an AI / ML capability, data science expertise is needed to build and iterate AI / ML models, analyze results, and ensure correct interpretation of the model decisions and reliability. Without significant attention to detail and understanding of the data source there is a risk of inaccurate models producing incorrect recommendations leading to undesired outcomes.

4.1.2 Classification

An important source of data in aviation today consists of safety reports that are largely composed of free-form, natural language. Incident reports are extremely valuable data sources that provide insights as to when, how, and why past events may have occurred along with the impacts of aviation system changes on safety outcomes. Unfortunately, analyzing and classifying safety reports manually is difficult and resource-intensive, which can leave valuable data unused and valuable insights undiscovered. To mitigate this issue, the standard approach refined over the last two decades has been to utilize natural language processing (NLP) machine learning models. This advancement has significantly increased the ability to extract meaningful insights from these reports supporting the classification and monitoring of changes to classified event sets over time.

A challenge to the consistent development and application of classification algorithms is that research surrounding NLP in aviation has focused on categorizing reports relying on labels given by domain experts without the use of a general taxonomy (Buselli, et al., 2022). Model bias can be introduced when using NLP to categorize data for inclusion in a training set, and in turn, this bias may unintentionally influence the decision of the model. Although efforts have been made to increase generality through the creation of structured data sets like Toolkit for Air Traffic Management (ATM) Occurrence Investigation (TOKAI) with some success, it is still a challenge to introduce new event categories without a significant level of effort to reevaluate legacy data sets (Buselli, et al., 2022) (Oneto, et al., 2021). This reinforces the need and potential benefit of developing common aviation taxonomies, as addressed in Section 5.3.

Another AI technique supporting data classification is the use of Large Learning Models (LLM), that can be used to process human communication language, such as that described in NAS safety incident reports or in pilot-communication voice recordings. While generative AI chatbots use either generally available or individually trained (e.g., proprietary) LLMs, there has been a recognized benefit to in developing smaller, focused, industry-specific data sets. Creating a specific LLM means that the training data can be verified for accuracy while reducing the training set to only what is necessary. In turn, these customized LLMs can be more reliable, efficient, and cost-effective making them more viable solutions for smaller organization; it is predicted that more and more organizations will adopt LLMs but through their own models (Snowflake, 2024).   LLMs, trained specifically for aviation data sets, can be applied to classify aviation event narratives.

4.1.3 Cluster / factor analysis

A significant advantage of artificial intelligence is the ability to further our understanding of data in ways that would be difficult without significant resources. While classification or categorization ML algorithms focus on automating tasks routinely performed by domain experts (e.g., determining a type of event), this type of ML application alone is "rarely able to support safety practitioners in the process of *investigation* of an incident after it happened" (Oneto, et al., 2021). Additional ML algorithms, including the use of topic modeling and clustering, are now being applied with promising results.

Topic modeling and cluster methods were used to automatically extract meaningful insights and identify recurrent trends from aircraft Loss of Separation reports as part of research sponsored by the Single European Sky ATM Research and Development programme (SESAR) (Buselli, et al., 2022). Research methods incorporated three distinct steps: Exploratory Data Analysis, a Syntactic Analysis algorithm to extract *Toolkit for ATM Occurrence Investigation* (TOKAI) taxonomy factors, and an Automatic Contribution Assessment model to assess if pilots, controllers, or both contributed directly to an incident. While this work focused on labeling contributing factors to an incident, the techniques have the potential for other types of incident factor identification. Specific ML techniques and outputs used within the aforementioned research are summarized below:

- Exploratory Data Analysis topic modeling: used to extract the most recurring topics from the data without requiring manual review. Once the topic modeling method was successfully executed, a clustering analysis was used to related discovered topics with the other incident factors of interest, e.g. causal factors.
- Syntactic Analysis: then used to automatically examine the text structure and meaning of the reports to connect them with the TOKAI categories: perception, memory, decision, action, conformance.
- Automatic Contribution Assessment model: used to validate the output of the Syntactic Analysis by predicting who contributed the most to the evaluated events.
    - In this research, both a Random Forest and Support Vector Machine algorithm were tested. Although deep learning approaches were evaluated and outperformed these shallow learning methods in many tasks, they were avoided due to the large amount of data necessary for training that was not available. Both (Oneto, et al., 2021) and (Buselli, et al., 2022) concluded that the ML models achieved sufficiently high accuracy to prove confidence in the syntactic analysis dependability.

The processes of partially (or fully) automating factor assessment of NAS events using ML has been demonstrated to significantly reduce the time required to identify factors as compared to similar efforts performed by human practitioners (Buselli, et al., 2022). Additionally, the application of ML reduces human bias in data pattern and factor identification.

4.1.4 Outlier Detection

Another application of ML is an outlier or anomaly detection which pinpoints unusual patterns in data which can reveal areas of emerging risk, areas for improvement within the NAS, and opportunities for proactive safety oversight action. While domain experts can study and identify data patterns, a process reliant on human review may miss complex anomalies involving multiple parameters that interact. As noted in *Anomaly Detection in Aviation Data Using Extreme Learning Machines,* “Aviation data is complex due to its high dimensionality, heterogeneity, multimodality, and temporality, making anomaly detection difficult” (Janakiraman & Nielsen, 2016). Machine learning can help to take on this complex task and through continued research, multiple ML solutions have been proposed and analyzed, including:

- Multiple Kernel based Anomaly Detection (MKAD): an advanced algorithm for detecting anomalies, for example in commercial aircraft data (Das, Matthews, & Lawrence, 2011). Notably, this algorithm has the ability to seamlessly handle heterogeneous data by creating individual kernels for each data source. Once the kernels are created, a one-class Support Vector Model is trained separating nominal and anomalous data (Janakiraman & Nielsen, 2016).
- Extreme Learning Machine model: a simple neural network model using fixed random numbers for input layer parameters. Its training solves a linear least squares problem converging much faster than other algorithms while achieving superior results according to Anomaly detection in aviation data using extreme learning machines (Janakiraman & Nielsen, 2016). Traditionally, these algorithms have been used in supervised learning applications where the anomaly detection is reframed as a classification problem. Initial indications are that the results match or exceed the results of MKAD algorithms and are trained up to 3 times faster; however, further research is needed on algorithm performance comparison.

Other studies have evaluated additional techniques for anomaly detection. These anomaly detection algorithms may have applications in identifying conditions that may be early indicators of impacts of regulation or oversight changes. Random Forest algorithms are an effective tree-based ensemble model that uses bootstrap aggregation and randomized node optimization (Oneto, et al., 2021). Decision Jungle algorithms were developed to counteract limitations found in Random Forest algorithms, namely in memory usage and overfitting as discussed in the study on *Machine Learning for Anomaly Detection and Process Phase Classification to Improve Safety and Maintenance Activities* (Quatrini, Contantino, & Patriarca, 2020). The study noted that tree and forest algorithms continuously grow in depth and complexity as more nodes are added. Decision Jungle algorithms that assemble decision-directed acyclic graphs instead of trees allow for multiple paths from the root to a leaf preserving crucial memory and avoiding overfitting. These ML algorithms were used to identify outliers within pharmaceutical data with anomaly detection in best-configured algorithms receiving over 99% precision. The collective set of outlier detection algorithms has the potential to be applied to aviation data sets to identify off-nominal or changing data patterns. As noted previously, this has the potential to identify emerging risk and support proactive oversight action.

4.1.5 Risk Assessment

Another recognized potential application of machine learning in safety management is the capability to support risk assessment. Machine learning can be applied to data sets to identify (and in the future recognize) data signatures or patterns that can be assessed for risk. This application has the potential to identify unknown hazards. With advancements in machine learning, previously unknown hazards can be identified along with associated risks to determine appropriate mitigation strategies.

An example of machine learning-supported risk assessment is proposed in the In-time Aviation Safety Management System Concept of Operations. This concept applies machine learning to support an efficient, real-time response to risk in all aspects of safety monitoring by leveraging airplane equipment data (Ellis, et al., 2022). Currently, equipment such as digital flight data recorders and cockpit voice recorders are required to be in place for Part 135 aircraft but are not required to flow data into a universal data store. If the aircraft equipment data were required to feed into data stores such as ASIAS, the data could be used for consistent predictive analysis, providing better insight into the productivity of current safety risk management procedures while identifying hidden or emergent risks in real-time. There are many identified applications of this approach in the vision for an In-time Aviation Safety Management System, including support for assuring safety in Advanced Air Mobility (AAM). A precaution noted in this concept is that any operational or regulatory change may disrupt current safety measures with potential impacts to flight safety. To properly introduce this approach without causing safety-critical mistakes, the concept of operations calls for significant research and testing before implementation.

4.2 Precursor and root cause analyses

In the context of system safety, precursors are broadly defined as events and conditions that preceded an accident or incident and are not necessarily causal. Precursors may occur in combination or in a specific sequence leading to an accident or incident. Perspectives on defining and applying precursors as part of a risk management regime vary. When considered in accident sequences, the National Aeronautics & Space Administration (NASA) and the Nuclear Regulatory Commission characterize precursors in terms of those initiating events and degraded conditions that could produce severe consequences, accounting for the probability and the extent of the failure when selecting candidate precursors for further analysis. Broader constructs that consider a precursor as any event or condition that could lead up to an accident or incident have been applied when seeking to promote workshop-style brainstorming as explored in a project to investigate accident precursor analysis and management, including use within the aviation industry among others (National Academy of Engineering, 2004).

Root causes can be considered as those events and conditions which directly result in an accident or undesired occurrence. Root cause analysis entails the systematic review and decomposition of an accident or undesired occurrence into the direct and contributing factors that caused that event. Those factors may span system failures, design flaws, process lapses, environmental conditions, and human factors, among others. Once root causes and contributing factors are identified, a closed-loop system for corrective actions is applied with the goal of preventing future occurrences. Within the U.S. SMS rule, the term "root cause" is synonymous with the term "hazard," which is defined as "a condition or an object that could foreseeably cause or contribute to an incident or aircraft accident" (14 CFR 5.5, 2023). This type of analysis is called for as part of U.S. Commercial Space Transportation licensing regulations in Part 437.73 and indirectly within the other Part 5 SMS rule provisions for hazard analysis. The UK CAA also provides guidance on *Effective Problem-Solving and Root Cause Identification* for airworthiness organizations with elements that address EASA's SMS provisions for root cause analysis and corrective action (UK CAA, 2019).

Methods for precursor and root cause analysis range from interview and diagramming techniques to deductive analysis methods as well as application of natural language processing and other artificial intelligence techniques to analyze incident data for potential causes. Root cause analysis can be used for examining regulatory oversight process lapses, and an understanding of root cause analysis processes and methods is needed when evaluating service provider or licensee approaches and artifacts for effective application of techniques to prevent future occurrences. The results of precursor analyses can be used to focus and prioritize oversight activities for surveillance and inspection. Examples of tools and techniques for precursor and root cause analysis include:

- **Five Whys** – An iterative process to facilitate interviews on why something occurred to drill down to lower sub-causes until the root is identified. Example applications of this and the next technique are presented in IATA training on root cause analysis for ICAO and EASA Safety Management System compliance (IATA, n.d.)
- **Fishbone (Ishikawa) Diagram** – A technique for eliciting causes according to categories for process, equipment, human, and environment factors, among others
- **Management Oversight and Risk Tree** – A comprehensive chart-based technique to identify management system and control inadequacies that need to be corrected to

prevent accident occurrence or recurrence. A top-level accident or undesired event is broken down into oversights and omissions or occurrences of previously accepted risk. Oversights and omissions result from the combination of less than adequate control factors for accidents and management systems which span policy, implementation, and risk assessment (Stephans, 2022).

- **Latent class analysis** – Identification of groups and distinct subgroups or categories in a data set through use of statistical techniques. An example application of latent class analysis for rejected take-off incidents as a runway accident precursor is presented in Oh, S., Kim, S. & Yoon, Y (2020).
- **Machine learning** – Use of unsupervised models or cluster analysis to identify potential precursors or contributing factors from large data sets of accidents and incidents and supervised models to classify and assess precursors to known safety events (see Section 4.1). An example application of machine learning for aviation incident precursor analysis is presented in a 2020 American Institute of Aeronautics and Astronautics conference session on using big data to increase safety (Ackley, 2020).

4.3 Barrier modeling

Barrier modeling (also known as bowtie modeling) provides a structured means to model risk, including threats that lead to a hazard, consequences, and mitigations or controls that either prevent the hazard or reduce the severity of consequences. Barrier models have features that make it suitable for understanding and visualizing risk, including (UK CAA, 2024):

- Organizing data and performing basic risk calculations
- Providing a structure and systematic means of analyzing and understanding hazards
- Facilitating an understanding of risk control measure suitability and weaknesses
- Supporting risk-based, data-driven decision-making
- Providing a visual tool to communicate risk management.

The models can be used to monitor and understand current control performance, identify the need for new controls, evaluate change, and support root cause analysis. These elements, shown in Figure 2 as reproduced from *Bowtie Methodology as a Meta-Analysis Tool for Aviation Accident Investigation Reports* provide the opportunity for at-a-glance understanding of key elements that drive, mitigate, and result from a hazard event (Kyriakos, 2020).

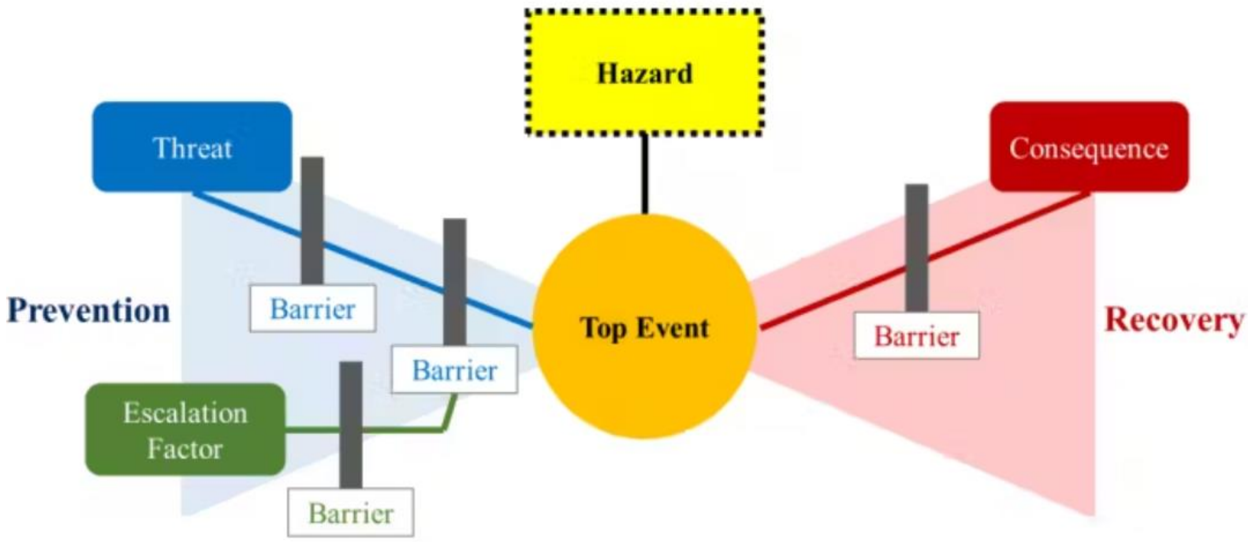


Figure 2. Bowtie element components (Kyriakos, 2020)

When building barrier or bowtie models, different methodologies can be used to perform threat analysis. For example, the FAA's Aeromedical organization developed a threat set considering conditions and contributing factors, downstream conditions that arise from an original condition, and adverse treatment impacts (Tvaryanas, 2024). When specific threats are identified, they are individually analyzed to determine barriers that would prevent the realization of an associated hazard event (considering technical, human, organizational, process, and environmental barriers), possible consequences, and any mitigations that would minimize exposure of the consequence. Systems analysis is applied to examine all system elements both to inform the model and to support the identification of goals, issues, and potential improvements within the system under review.

A case study in the use of bowtie models for incident and hazard (root cause) analysis is provided in Australia's analysis of helicopter collisions with terrain where tail rotor pedal failure is involved (Australian Transport Safety Bureau, 2020). In this case, a bowtie was created to model the realized event and threats identified as contributing to the event along with associated barriers. Here, a barrier model helps to identify the likelihood of specific threat occurrence and to evaluate realized barrier effectiveness (i.e., identify when a barrier worked or did not work). Other bowtie models developed by the Australian Civil Aviation Safety Authority are organized into risk categories including loss of control inflight, controlled flight into terrain, mid-air collision, runway incursion/excursion, human performance, and organizational factors (Australia Civil Aviation Safety Authority, n.d.).

Another application of bowtie models by an ANSP is demonstrated by the United Kingdom Civil Aviation Administration (CAA). This organization performs barrier-based audits (including UK Safety and Airspace Regulation Group audits and internal audits) to the barriers or controls identified within the bowties (UK CAA, 2024). Their work helps to evaluate barrier performance, identify risk vulnerabilities; inform business risks and organizational challenges; and provide data for proactive oversight to address areas of vulnerability. Barriers and associated indicators include both activity-based measures and outcome-based measures, including failures/accidents, precursors, and results (UK CAA, 2024).

Bowtie models are also applied within aviation industry systems given applications for safety risk management workflows using a structured and systematic approach to capture and manage risk. For example, Boeing has used barrier modeling tools to capture hazards, outcomes, and barrier data as an integrated part of risk management and safety assurance (Escobar, 2024). Barrier performance assessment also provides data to inform aviation maintenance decisions including replacing parts, increased frequency of maintenance, updated maintenance manuals, new designs, and additional process checks.

4.4 Occurrence analysis

Occurrence analysis is a capability that augments *hazard analysis* to provide insights that are not available through hazard analysis alone. The National Transportation Safety Board (NTSB) has 96 discrete occurrences that investigators may select from to answer the question, "What did the hazard cause?" If there is an occurrence of interest, such as a wrong-surface landing, a runway excursion, or a mid-air collision, automation tools can quickly determine which hazards have been cited as causal to those events in the past. Not only can the hazards be identified, but their relative frequency can be assessed to provide insight into which hazards are most often causal to the occurrence. This capability can then be expanded to not only identify individual hazard citations but also the most frequent combinations of hazards that have caused the occurrence.

Finally, just as with hazard analysis, the severity of each occurrence can be determined to provide a probability distribution for potential outcomes. To illustrate this approach, if analysts are considering hazards that have caused mid-air collisions in the past, data can be used to assess the potential severity of a mid-air collision. As an oversight tool, this capability would allow the overseer to evaluate whether a risk analysis contains sufficient information to capture the true nature of the risk. As a routine application, risk assessments that are allowed by policy to base risk findings on a single severity outcome could be compared to a more robust assessment that considers the probability of effects at all severity levels. As an example, a typical risk assessment of a hazard that could cause a midair collision would generally rely on a severity level of "catastrophic." As an example of the results an overseer might produce as part of an occurrence analysis, the NTSB data for the time period between 2008 and 2022 is provided in Table 4.

Table 4: Mid-Air Collision Severity Data, 2008-2022

Occurrence	Total	Catastrophic	Hazardous	Major	Minor	Minimal
Mid-Air Collision	249	72	56	86	30	5

This data shows that while catastrophic outcomes are possible, about 71% of midair collisions resulted in a lower severity outcome. This might indicate to an overseer that additional investigation is needed to determine the risk associated with multiple levels of severity.

4.5 Sequential / simultaneous hazard modeling (influence diagrams)

Hazard analysis was pioneered by the US Army Air Corps in the wake of World War II and refined during the 1950s by NASA in collaboration with industry partners as the nation prepared for manned space flight. Compared to today's aviation environment, aircraft, and spacecraft in the 1940s and 1950s were relatively simple machines, and accidents were frequently caused by the occurrence of a single hazard. Today, any random sampling of aircraft accident reports will reveal that accidents are only caused by a single hazard in rare cases. Instead, a combination of hazards occurring either at the same time or in a short sequence is usually required to cause an accident. While the aviation environment has matured in terms of both complexity and capability, many aviation safety management systems continue to rely on the 75-year-old technique of assessing risk on a hazard-by-hazard basis and lack mathematical capabilities to account for higher accident probabilities when multiple hazards occur.

To address this shortfall, model-based approaches have emerged to quantify not only the risk of a single hazard occurring but also any combination of hazards. This is done by representing the uncertainty associated with hazards, along with any dependencies that may exist, in an influence diagram. That diagram then becomes the blueprint for an analytic model such as a risk framework or a decision tree.

To illustrate the model-based approach, consider two standardized hazards within the NTSB and AHCS taxonomies: *main landing gear failure* and *crosswind*. On every landing, there is some probability of experiencing a crosswind, and also some probability of a main landing gear failure. These uncertainties may be represented in an influence diagram as ovals as shown in Figure 3.

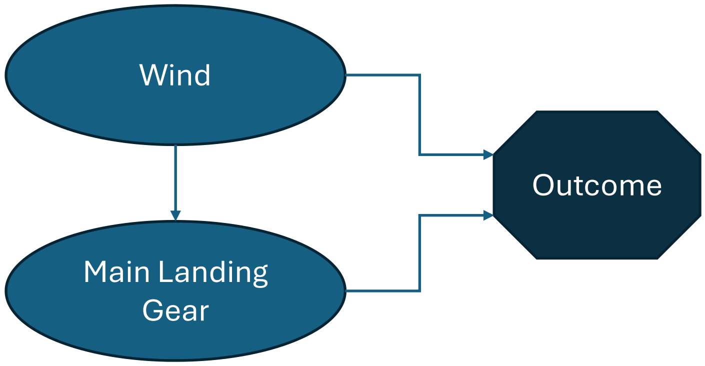


Figure 3: Influence Diagram for 2 Hazards

The arrows represent dependencies. For example, the arrow from "Wind" to "Main Landing Gear" indicates that there is some probability that the main landing gear will fail when there is no crosswind, and a different probability when there is a crosswind. The influence diagram can be converted to a risk framework as shown in Figure 4.

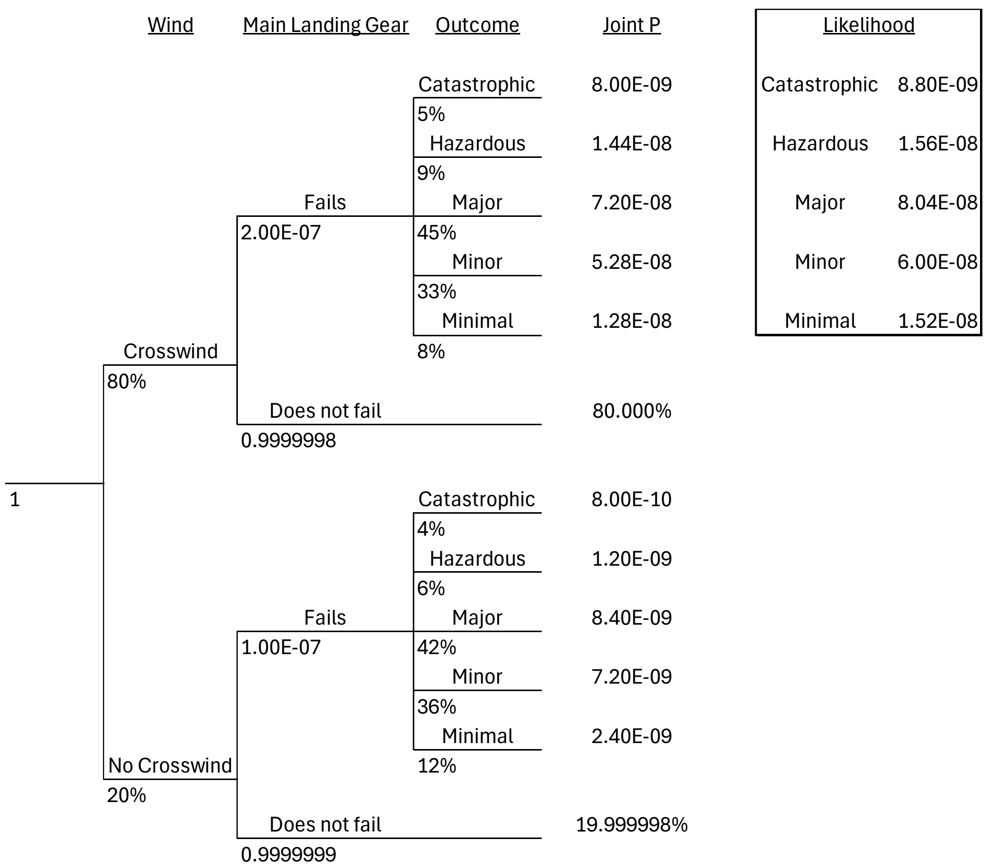


Figure 4: Risk Framework for 2 Hazards

The risk framework could support an overseer by verifying the data used for the assessment is appropriate, then providing independent confirmation of the results. Note that the notional data in Figure 4 indicates that a main landing gear failure is twice as likely to occur with a crosswind than without a crosswind. Also, the *likelihood* of an effect at any severity level is determined by summing together the joint probabilities of outcomes at each level of severity.

4.5.1 Decision trees

This approach can be expanded to inform oversight decisions by regulatory authorities. Adding to the previous example, suppose the analysis is for a single-runway airport that is applying for certification. An alternative configuration could be to add an intersecting runway, allowing the airport to shift operations to a runway that is more closely aligned with the prevailing winds. This would reduce the probability of aircraft operating with crosswinds, perhaps to 10%. As a regulator, a decision would have to be made to either approve the single runway configuration or reject it, requiring development of the intersecting runways. This decision is reflected in the updated diagram in Figure 5.

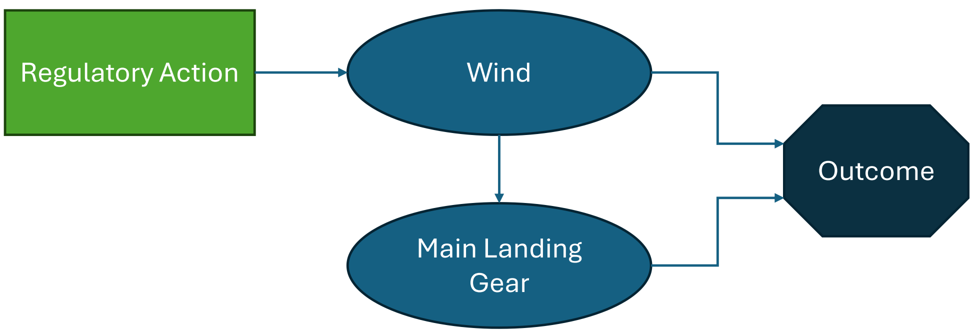


Figure 5: Decision Diagram for Regulatory Decision

This diagram serves as the blueprint for the decision tree shown in Figure 6.

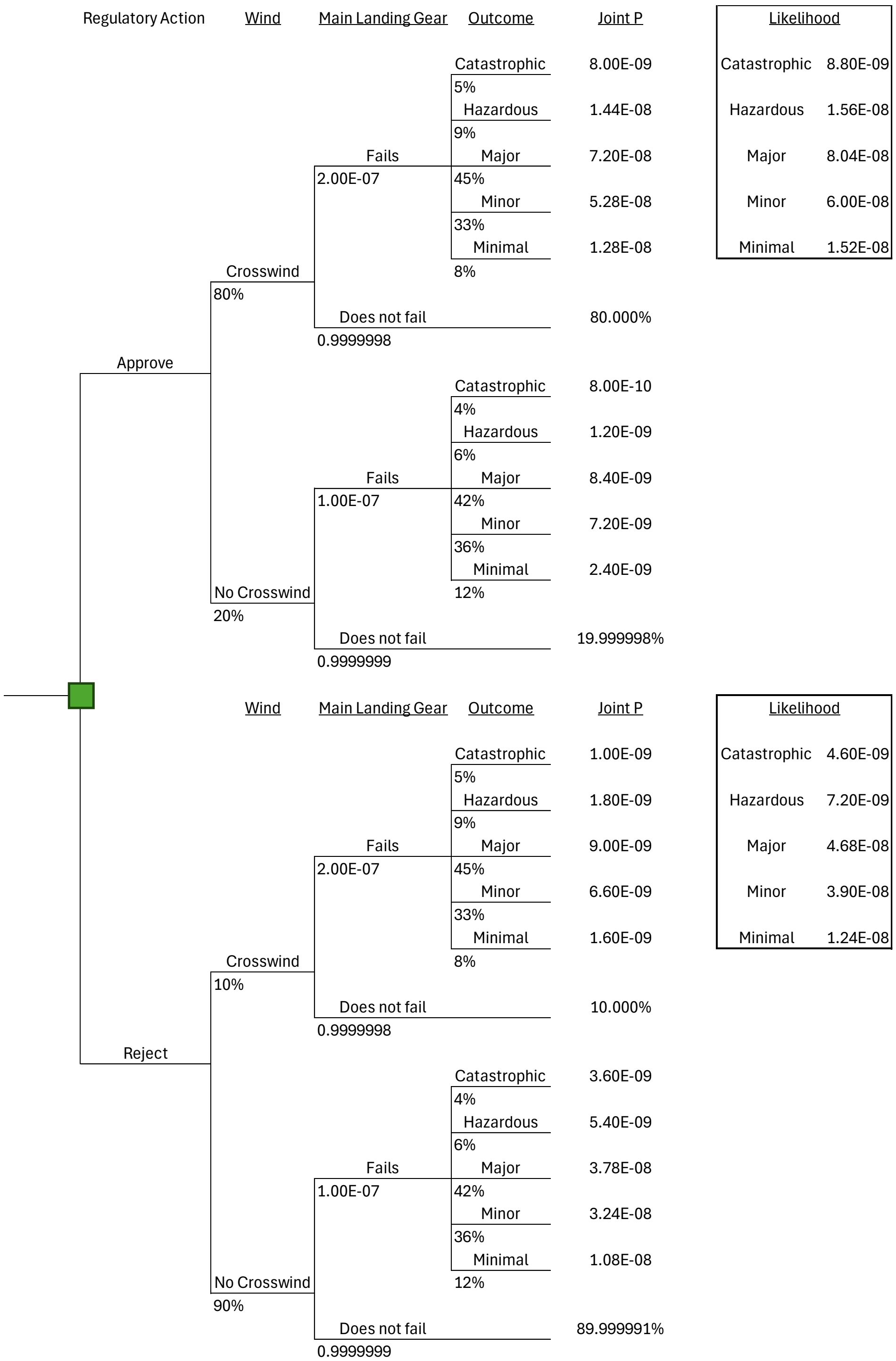


Figure 6: Decision Tree for Regulatory Decision

The risk framework and the results it produces can be automated to alleviate the analyst from having to manually perform the type of calculations depicted in Figure 4. Value data can also be added to the decision tree to assist with evaluating cost versus expected benefits. Automating these widely used and well-vetted processes establishes a foundation from which additional capabilities such as simulation, machine learning, and artificial intelligence become possible.

5 Safety information management

Information management systems and tools should accommodate different types of data needed for oversight from strategic planning throughout certification and continuous safety assurance. Data used to draw safety insights and support oversight decisions and actions should be credible, comprehensive, clear, current, and relevant. Data analytics, including those associated with "big data," depend on consistency within the data to allow machines to identify patterns or trends. To this end, the more the information can be standardized, the more powerful its potential as a tool for generating meaningful safety insights.

A data and information lifecycle, adapted from (Healey, 2024) is shown in Figure 7. This section addresses tools, technologies and techniques that support one or more areas of this lifecycle.

Figure 7. Data / information lifecycle

5.1 Big data

According to IEEE, Big Data "entails collecting, storing, processing, and analyzing immense quantities of data that is diverse in structure in order to produce insights that are actionable and value-added" (IEEE, 2024). In the context of aviation, ICAO notes that the concept of Big Data is interchangeable with business intelligence and data mining, where many industries are seeing the expansion of available data due in part to (ICAO, 2021):

- A reduction in the cost of collecting, storing, and processing data
- New sources of data and improved access to both existing and new sources
- New tools and techniques that leverage big data
- A broader opportunity for utilization of collected data

With access to more data, there are also opportunities to leverage data to achieve desired performance including safety outcomes. As noted in an article on *Big Data and [a] Smart Aviation Information Management System*, "because big data can provide multi-dimensional, adequate, and real-time information, intelligence can effectively translate this information into knowledge and capability, improving the predictive and preventive capabilities of aviation flight risks" (Dou & Tan, 2020). Within the aviation industry in general, Big Data can help to inform and transform individual aviation sectors such as the manufacturing sector. Additionally, intelligence information technology that is built on a foundation of Big Data has opportunities to enhance predictive perspectives and inform action to mitigate or prevent risk in the aviation industry. To take advantage of the potential of Big Data, there is a desire to better understand information models and information paths. Some researchers have begun to leverage information system modeling tools to capture and understand data relationships and to perform analysis (e.g. spectrum and coupling analysis) which examines the frequency of individual information nodes and characterizes the relationship between specific nodes. Also noted is a desire to improve cooperation and information-sharing mechanisms when it comes to aviation data. To do this, it is argued that the development and implementation of standards for data and data platforms will be an important issue. Here there may be a role for both industry and regulators (e.g., FAA) to drive or inform this process.

Another new utility for Big Data is as a transformative tool for developing measures statistics. This has been recognized in the aviation industry to improve accuracy and reduce costs associated with statistic development (ICAO, 2021). As Big Data is embraced and applied within regulated organizations, for example, airline operators that are exploring the integration of real-time aircraft monitoring with predictive maintenance platforms (Odarchenko, Hassan, & Zaman, 2019), this could also support access to and application of additional data by safety oversight organizations with appropriate agreements in place. This concept can be extended to where Big Data in the aviation domain may improve connectivity between aviation system participants, promoting access to and application of large data sets to understand safety performance and changes. Since the aviation industry operates within legal and regulatory frameworks, Big Data solutions will need to operate within data regulations that stimulate advancement and collaboration. Additionally, quality data is essentially Big Data, and due to the large number of stakeholders in this domain, quality and privacy-focused protocols for data exchange and usage will be needed (Aerotime Advisory Team, 2023).

With the emergence of large data sets and motivations to leverage the data, there are emerging and scalable tools and techniques to collect, manage, and apply Big Data. Some of these tools have been developed in the context of software frameworks supporting distributed collection, management, and processing of data across computing clusters, for example, Hadoop. Hadoop offers various components including (Iswarappa, 2015):

- Hadoop Distributed File System: a file system that provides mechanisms to efficiently store and access large amounts of data
- Hadoop YARN/Map Reduce: provide computational job scheduling and resource management
- HBase: defines a scalable and distributed non-relational database supporting fault tolerate storage and fast data search and access
- Pig: a data flow scripting language supporting
- Hive: a data warehouse infrastructure to support management, query, and analysis of large data sets
- Avro: a data serialization system that converts data into a compact binary format, where Java Script Object Notation (JSON) is used to define data types and protocols
- Spark: a compute engine for Hadoop data supporting stream process, machine learning, and data processing engines
- Flume: a service for aggregating and moving data

Additional tools are available or are being developed to support elements of Big Data. These include workflow management platforms (e.g., Airflow); open format data storage layers to facilitate access to and security of big data stores (e.g., Delta lake); query engines to search data in different formats (e.g., Drill, Presto); stream processing frameworks to support computations across data streams (e.g., Flink); distributed event streaming platforms (e.g., Kafka) that provide high-performance data pipelines, streaming analytics and data integration; and big data processing platforms that provides computer clusters and work with data lakes to support rapid development and data exploration.

5.2 Blockchain technology

The National Institute of Standards and Technology provides an overview of blockchain technology described as digital ledgers or shared records in which users record transactions with secure mechanisms that prevent recorded transactions from being changed (Yaga, Mell, Roby, & Scarfone, 2018). At its core, block chain is a decentralized, encrypted, and distributed data repository (Ahmadisheykhsarmast, Aminbakhsh, Sonmez, & Uysal, 2023). Participants submit transactions to a node within a blockchain network where it is propagated to other nodes in the network; a pending transaction may wait in a queue for it to be executed or published, which creates a block that includes a header with metadata supporting validation and authentication and block data which describes the transaction and ledger events within the block (Yaga, Mell, Roby, & Scarfone, 2018).

Key concepts within blockchains include:

- Ledger: technology that maintains records and provides full transactional history
- Security: application of cryptographical technology to ensure data within the ledger is not tampered with
- Shared access: the ledger is shared amongst participants with transparency across the blockchain network; while shared blockchain can be "permissioned" where access is limited to authorized persons
- Distribution: The blockchain implementation can be distributed supporting scaling and resiliency

As part of the blockchain technology overview, a decision flow diagram, initially developed by the US Department of Homeland Security, is provided to help evaluate blockchain suitability for general information management needs as shown in Figure 8.

Blockchain technology has applicability for use cases beyond its origins in cryptocurrency and financial instruments, for example, those transactions requiring multi-party collaboration with needs for storing and sharing sensitive information, a single source of truth where transactions are secure and immutable, and software automation and enforcement of communication protocols and processes (Wang & Liu, 2023). As safety oversight often requires a shared understanding of data with inputs from multiple participants, blockchain technology can be considered in the design of new and evolving oversight data sharing and management platforms.

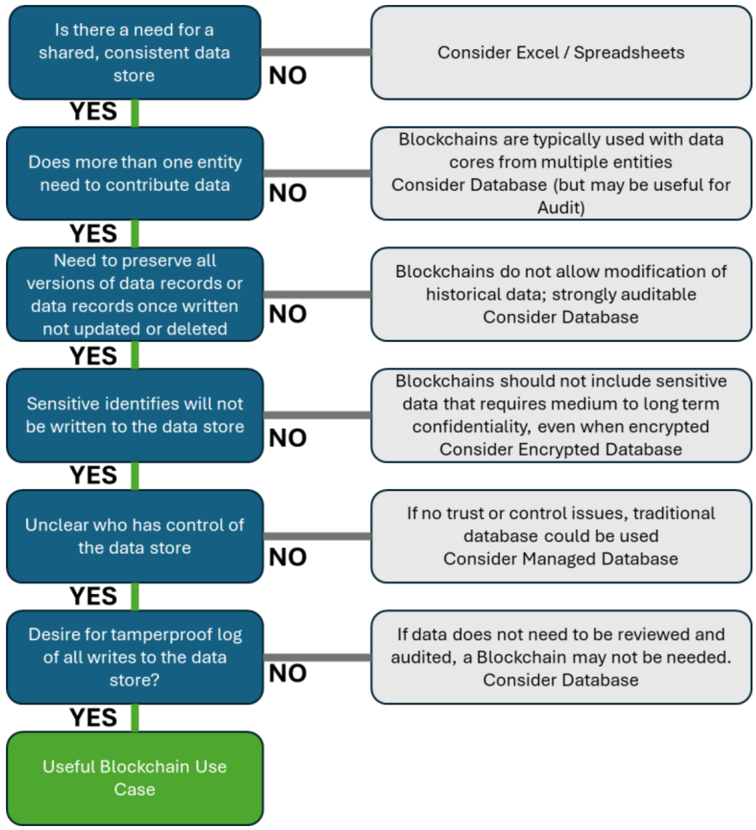


Figure 8. Identifying blockchain use cases (reproduced from (Yaga, Mell, Roby, & Scarfone, 2018)

Within the aviation industry, EASA sponsored a case study from 2022 – 2024 on the potential use of blockchain technology to manage aircraft component certifications from production approval throughout that component's lifecycle as changes are periodically introduced. Some of the key objectives of the case study were to evaluate blockchain solutions to manage aircraft part and component certification and maintenance data and to address regulations and standards needed for blockchain implementation (EASA, 2024a). Using blockchains for traceability in parts' certification data and secure tracking, the EASA study examined cases for validating the authenticity of issued certifications and managing certificate revocations given known parts' defects as well as suspected unapproved parts.

EASA's study also referenced additional use cases for blockchain within the international aviation community (EASA, 2024b). Japan-based JALUX Group invested in Block Aero blockchain technology in 2019 to transform quality management data for aircraft and engines from paper-based documentation to secure digital data (JALUX, 2022). In 2024, Air France Industries KLM Engineering and Maintenance group in partnership with Parker Aerospace announced the successful implementation of SkyThread blockchain technology to manage aircraft parts traceability and tracking from design through installation, repair, and removal on the 787 fleet (Air France Industries, 2024).

Honeywell is also using blockchain as part of its aircraft parts marketplace to ensure parts listed for sale include certification data, accurate images, and quality records along with a digital ledger accounting for all part-specific transactions and repairs (Honeywell , 2020) (Honeywell, 2024). Missing or incomplete records in aircraft parts' repair data and certification records not only impact parts' resale value but also have implications for system safety. The potential use of blockchain to securely maintain and verify the providence of aircraft part maintenance and repair records is explored further in a 2022 IEEE publication on *Blockchain Information Based Systems in Aviation: The Advantages for Aircraft Records Management* (L. F. F. M. Santos, 2022).

5.3 Safety taxonomies

Safety taxonomies help standardize and organize aviation safety data for oversight review and analysis. For example, the Aerospace Hazard Classification System (AHCS), developed under an FAA cross-organization effort, is a system that translates multiple aviation safety taxonomies into a standardized format. One purpose of AHCS is to offer a standard taxonomy for adoption by any organization that does not already have one. For organizations with an existing taxonomy, AHCS acts as a translator to map terms to a common standard. In an oversight application, AHCS can support oversight initiatives to evaluate the effectiveness of a safety management system by mapping non-standardized hazards to its common standard which in turn allows the use of safety data to measure the influence of mitigation strategies on prevailing accident and incident rates.

AHCS relies on Systems Modeling Language (SysML) to define its terms and the relationships between terms. This also allows AHCS to be modular and available for import into other SysML models, such as the FAA's National Airspace System Enterprise Architecture in the emerging Unified Architecture Framework. The figure below provides a SysML Block Definition Diagram of the AHCS components (FAA, 2024).

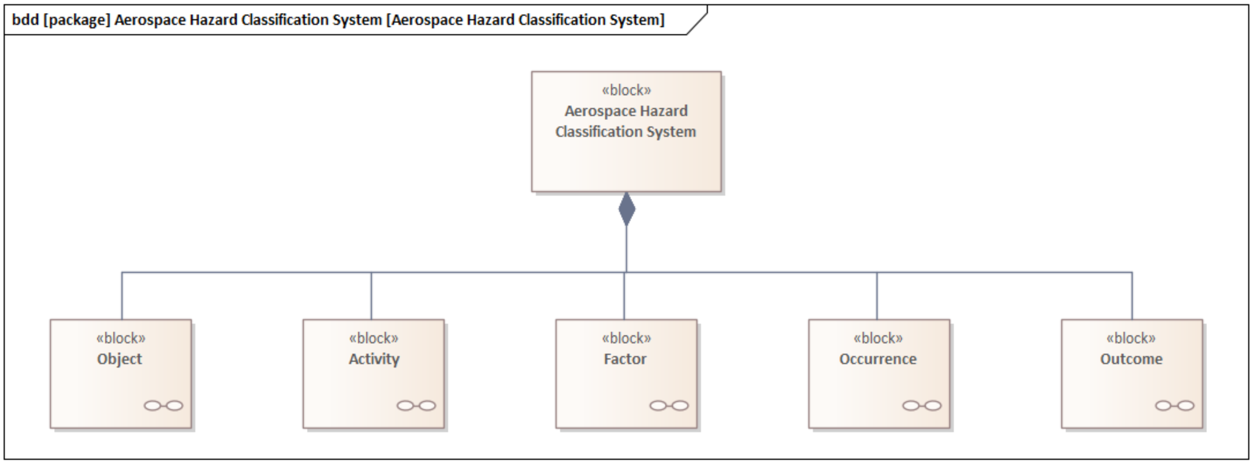


Figure 9: AHCS Block Definition Diagram

As an example application of AHCS, many accidents have been caused by pilots flying when impaired (e.g., after consuming alcohol). As part of an aeromedical safety risk assessment, one analyst may label this hazard as "flying while intoxicated," while another might label it as "flying under the influence" with many other hazard naming possibilities as varied as the analysts and organizations involved. AHCS maps all of these terms to a single, standardized hazard, "impairment/incapacitation due to alcohol."

The value of standardizing hazards becomes apparent when comparing the objectives of accident investigation to those of risk analysis. Figure 10 illustrates the dependent nature of this relationship.

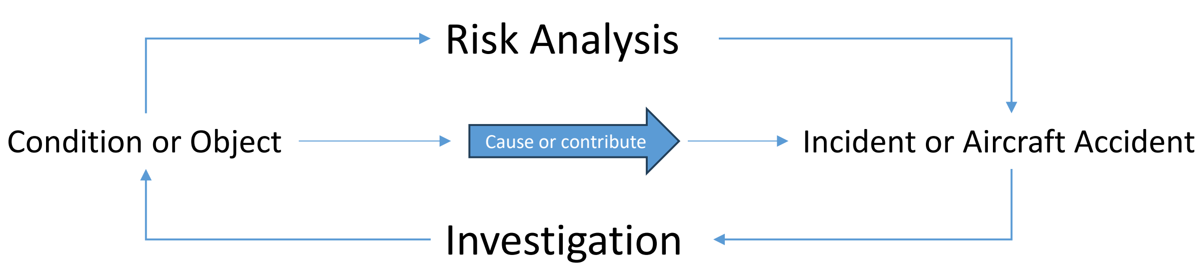


Figure 10: Accident Investigation & Risk Analysis

AHCS begins by adopting the definition of a hazard as set forth in Federal Aviation Regulations, "a condition or an object that could foreseeably cause or contribute to an incident or aircraft accident." During accident investigations, the NTSB identifies the hazards that caused or contributed to the incident or accident. This is done using a standard taxonomy that allows every investigator to select the same label every time the hazard is identified as a finding in an investigation. The hazard described above would be cited as "NTSB Code 02012015xx, Personnel issues – Physical – Impairment/incapacitation – Alcohol." This coded hazard demonstrates four levels of decomposition, with the final two digits reserved for modifiers that add specificity. The hazard is further described in a text-based causal narrative within the report.

During risk analysis, safety professionals estimate the likelihood of a hazard causing or contributing to an incident or accident. Historically, this has been done through qualitative methods and without the benefit of a standardized hazard taxonomy. Therefore, if analysts identify a hazard such as “flying while intoxicated,” the ability to cross-reference the applicable NTSB finding data for accident causal factors is impeded (especially for data analysis tools without a human in the loop). By standardizing the hazard, AHCS unlocks the opportunity to use data to quantify the risk associated with any hazard. This data includes the number of times the hazard has been cited in the past along with the severity of the outcome in each case.

As a translator for other aviation taxonomies, AHCS maps to certain recognized taxonomies, including the aforementioned NTSB accident investigation classification system. Additional taxonomies supported by AHCS include the FAA’s Accident Incident Data System (AIDS) codes, CAST/ICAO Common Taxonomy Team (CICTT), NASA’s Aviation Safety Reporting System (ASRS) coding, and FAA Mandatory Occurrence Report (MOR) classifications among others.

5.4 Data exchange models

Information management systems are often siloed, preventing the interoperability and benefits that could be gained from the harmonization and integration of multiple aviation data sources. Historically, aviation information exchange has required significant time and effort from individual organizations due to their incompatibility with international system-to-system data sharing. This makes the addition of new concepts and data elements challenging (ICAO, 2022).

Information exchange models standardize the way data is formatted between organizations and systems allowing for more efficient exchange of information. Current exchange models enable the sharing of aeronautical, flight, and weather information machine-to-machine. As more users adopt common exchange models, the greater the benefits of their implementation will become to the aviation community (ICAO, 2022). Examples of aviation data exchange models include:

- **Aviation Information Data Exchange (AIDX) -** The global XML messaging standard for Airport Collaborative Decision Making supported by ICAO for exchanging flight data between airlines, airports, and any third party consuming operational data (IATA, n.d.).
- **Incident Data Exchange (IDX) -** A worldwide, aggregated, de-identified database of incident reports including flight operations, cabin, ground operations safety, and security

occurrences in accordance with ICAO requirements for Safety Management Systems (IATA, n.d. 2)

- **Airline Industry Data Model (AIDM) -** A single point of access to store industry-agreed vocabulary, data definitions and their relationships, and related business requirements maintained by the IATA's Architecture and Technology Strategy Board (IATA, n.d. 3)
- **Aeronautical Information Exchange Model (AIXM) -** Enables the collection, verification, dissemination, and transformation of digital aeronautical data including aerodromes/heliports, airspace structures, points, NAVAIDS, procedures, routes, and flying restrictions. Maintained by the FAA's and EUROCONTROL's AIXM Change Control Board to enable states to comply with the ICAO global and regional requirements for the provision of aeronautical information (EUROCONTROL, n.d.)
- **Flight Information Exchange Model (FIXM)** - A global exchange standard capturing Flight information implemented in Unified Modeling Language and XML to support the data exchange requirements for the *Flight and Flow Information for a Collaborative Environment* concept as defined by ICAO's Air Traffic Management Requirements and Performance Panel (ICAO, 2022).

5.5 Data visualizations and dashboards

Dashboard analytics tools that combine performance indicators with data visualization can help to manage performance and facilitate data-driven decision-making. Charts, maps, tables, and other infographics are popular visualization techniques for dashboards, though care needs to be taken to match visual information to user needs and the actions and conclusions to be drawn from the information (Kobi, 2024). In addition, an important check for dashboards is conducting usability tests to validate the achievement of the intended purpose and to address human factors for information processing load. Traditional business intelligence logic tools such as Microsoft PowerBI, IBM Cognos Analytics, and SalesForce Tableau provide frameworks to design a dashboards and data visualizations as well as capabilities for data cleaning, query building, and report publishing. However, the incorporation of usability in data visualizations and dashboards requires determining dashboard needs and requirements that shape the data views, filtering, and customization capabilities.

As part of European Union research into aviation safety dashboard best practices (Future Sky Safe, 2019), a dozen Air Navigation Service Providers were engaged in an effort to prototype executive-level safety dashboards and produce design guidance, including:

- Thinking in terms of topics of interest vs. specific indicators
- Addressing operational safety, people and culture, the technical system, and change management
- Balancing between reactive and proactive indicators
- Developing agile dashboards that can be evolved
- Incorporating user-centered design approaches

The same study also produced a thorough list of common pitfalls in safety dashboards, including:

- Always green: Indicators whose status is always reported as "good" and could drive complacency or a false sense of positive performance.
- Impulsive reactions: When a status turns "red," quick action is assumed to be needed, even though the most appropriate and effective resolutions benefit from a full understanding of causes and contributing factors.
- "Targetology:" Too much focus on meeting a defined target may suppress reporting or prompt changing analyses to meet targets.
- Frozen indicators: Indicators that are not defined at the right level may be insensitive to changes of interest.
- Perspective: Safety events need to be understood in relation to other factors (e.g., normalized for operations counts).
- A new normal: Organizational constraints should not drive the normalization of risk, including reducing the data collected and applied within dashboards.
- Quantity vs. quality: More indicators is not necessarily better, rather value comes in assessing effectiveness of safety case or safety activities.
- Black swans: Masking of individual events or reports that do not appear to make sense or reflect nominal operations may prevent the opportunity for proactive action.
- Data silos: Relationships between different information on dashboards may be difficult to realize; engaging a wide range of stakeholders in dashboard development can help to address this.
- Hyper-focus on Top-5: Focus on headline-related indicators may mask opportunities to monitor and address "smaller" (i.e., less consequential) risks that may lead to "bigger" risks.
- Success only: Many organizational cultures have pressure to report only good news.
- Reactive focus: Measures that focus only on realized events, missing opportunities to define forward-focused measures and evaluate potential safety impacts of changes.

6 Safety objectives and performance

Aviation safety objectives are essential to safety risk management or safety oversight as they define the overarching system goals. At their highest level, safety objectives should focus on the degree of harm to people and the degree of loss of property. Once safety objectives are established, data should be used to measure the organization's performance in terms of achieving the objectives. The following sections describe safety objectives in more detail and then describe the process of using data to measure performance.

6.1 Safety objectives

ICAO defines a safety objective from the regulator's perspective as a "high-level statement of safety achievement or desired outcome to be accomplished by the State Safety Programme or service provider's safety management system." Safety objectives are intended to support the development of measurable targets against which to assess performance. Although this concept is similar to the discussion of performance indicators in Section **Error! Reference source not found.**, a distinction is the topic and scope of safety objectives. Safety objectives are safety priorities of an organization defined to recognize and address key elements of safety risk.

In establishing specific safety objectives, ICAO provides guidance on leveraging safety planning activities to drive the definition of safety objectives, a minimum level of safety performance is used to drive objectives and subsequent performance measures. (ICAO 9859, 2018). When establishing safety objectives, there is a general understanding that the objectives align with safety goals associated with a safety oversight system, inclusive of safety risk management, safety assurance, safety promotion and safety policy. Therefore, goals in these must be defined addressing elements including (Britten, 2022):

- Reporting policies / requirements
- Management / employee responsibilities
- Privacy and security of reported incidents
- Issue management
- Safety communication and data sharing
- Management / employee relationships

Defined goals should be representative of organizational purpose and strategy. They collectively outline a roadmap to identify and manage risk, gain employee buy-in, boost team performance, and drive execution of strategy (Han, 2023). To ensure the collective set of objectives fulfills the organization purpose and strategy, a strategy map that, models cause-and-effect relationships, may be useful (Han, 2023).

A risk-based approach to defining safety objectives for a regulator is shown in an example from Australia's Civil Aviation Authority (CAA). Australia's CAA's National Aviation Safety Plan lays out national aviation safety risks and challenges and serves as a foundation to derive safety objectives (Australian Government, 2021-2023). The safety objectives "represent the desired outcome Australia seeks to achieve to address its identified aviation safety risks and enhance overall safety within the aviation sector and to the travelling Australian public" and are as follows:

- Goal 1: Improve the safety of Australian aviation operations across all sectors where Australia has State oversight responsibility
- Goal 2: Strengthen Australia's safety oversight capabilities
- Goal 3: Embed an effective State Safety Programme that delivers an acceptable level of safety performance
- Goal 4: Reduce the likelihood of Australians being involved in an aviation accident outside of Australia by supporting and influencing global aviation safety
- Goal 5: Expand the use of Industry Safety Programmes by Australian Industry
- Goal 6: Ensure Australia has the appropriate aviation infrastructure to support safe operations

In a further hierarchical fashion, each safety objective is mapped to one or more safety performance measures and targets.

Other approaches for specifying safety objectives are informed by rule design and the type of regulations in place to facilitate the measurement of safety performance. A compliance-based regulatory strategy relies on formalism and rigid enforcement to drive greater compliance. However, this strategy may not inspire a cooperative relationship between the inspector and the licensee. Performance-based regulation, focusing on outcomes to be achieved, is often reactively focused on outcomes and with a tendency to give rise to narrow safety assessments (Bernard, 2024). Therefore, in considering safety objectives as a measure of *overall* safety performance, there is need to ensure objectives are not simply limited to outcomes. Rather, objectives should address a range of facets of the safety management system, including not only outcomes but also performance of organization, process, and culture. This concept aligns with ICAO's framework for safety performance measurement discussed above.

One approach to establishing safety objectives can be adapted from rule design. A framework for rule design adopted from rule design for high-risk industries (Coglianese, 2024) summarizes an approach in four general categories, as shown in Table 5.

Table 5. Rule design approach categories

	Means	**Ends**
Micro	Micro-based rules (“prescriptive regulation” on how a requirement must be achieved)	Micro-based rules (“performance-based regulation” such as noise limits, emission standards, and testing protocols)
Macro	Macro-means rules (“management-based regulation or management systems”)	Macro-end rules (“after the fact liability-based regulation” such as penalties for misconduct)

In the table above, *means-based rules* drive regulated organizations to take or avoid actions while *ends-based* are mandates to achieve or avoid specific outcomes; in this case the distinction is that macro means tends to focus on the ultimate action, outcome, or problem where micro focuses on the individual contributors in the causal pathway to an action, outcome, or problem (Coglianese, 2024).

A similar framework can be applied to the definition of safety objectives, but when specific objectives are defined, the advantages and disadvantages of individual categories should be considered. For example, micro-mean rules are generally applied when exact design, behaviors, or actions are desired, not providing flexibility for a regulated entity but rather inserting tight control of a requirement. At the safety oversight organizational level, this type of specificity and associated prescriptive-based objective would be unusual. An approach with more flexibility in defining specific design, behavior, or action is in the application of micro-based end objectives. Here, a performance-based objective is defined, providing flexibility in how it would be achieved. This approach supports innovation and evolution of processes but also underscores the importance of effective compliance and performance monitoring to ensure that data is complete and sufficient and that regulated entities are adhering to the overall intent of performance-based objectives.

Macro-means objectives would measure the performance of a management system, including management practices and outcomes. This would include objectives that would demonstrate the capability to perform effective hazard analysis and identification of critical control points and demonstrated capability to address and reduce risk. To ensure objectives are truly met, evidence and data provided for these types of safety objectives would need to demonstrate meaningful and robust planning and analysis (Coglianese, 2024). Finally, under the concept of micro-based rules, actions would be motivated through punitive outcomes for objectives not met. Similar to micro-based means objectives, it is not expected that this would be a concept used for establishing safety objectives.

For rulemaking, there are times when regulators may choose to combine rules of different designs to address different elements of an oversight space and different types of regulated entities; however, when combining rules, work is needed to ensure different rules avoid working at cross-purposes and do not require cost without delivering corresponding benefit (Coglianese, 2024). In a similar nature, if applying different categories of safety objectives, the resultant set – whether addressing sets of macro-means objectives or macro-means objectives together with micro-based end objectives – should be validated to ensure they drive to a common purpose and expected benefit.

When evaluating safety objectives in the context of assessing an overall oversight program, especially those objectives that are not explicitly outcome-focused but instead related to desired outcomes specific to a management system (e.g. macro-means objectives), inputs of participants and stakeholders of the management system objectives are valuable data points. The collection of data from participants is often achieved through surveys and observation. A well-defined set of questions applied to the implemented data detection means can result in a data set that can be analyzed with descriptive and inferential statistics, including if specific safety oversight program elements had an impact on the safety objective.

6.2 Safety performance indicators

According to ICAO, safety objectives "should be taken into consideration during subsequent development of safety performance indicators and targets." ICAO goes further to provide a framework, reproduced in Figure 11, for safety performance measurement in terms of outcomes, service provider behaviors, and regulator or oversight organization behaviors (ICAO SM ICG, 2019a).

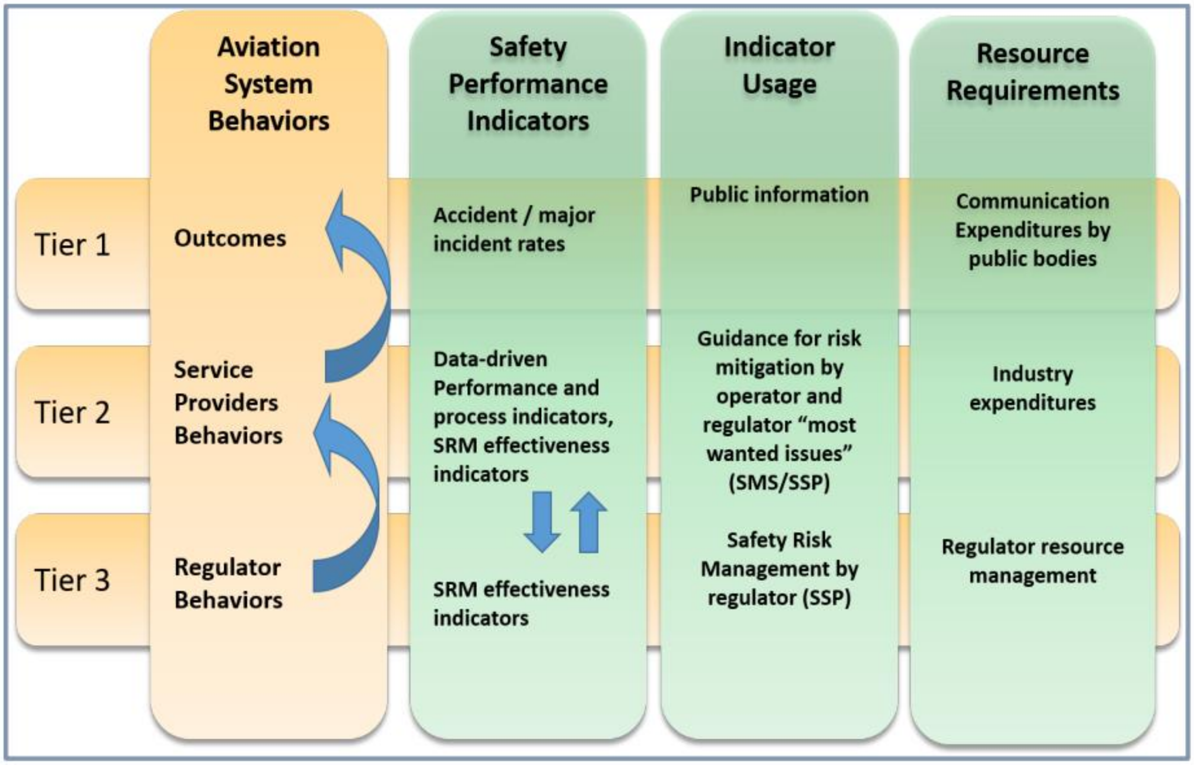


Figure 11. SM ICG framework for safety performance measurement (ICAO SM ICG, 2019a)

The purpose of this multi-tiered framework is to ensure that not only data focused on outcomes is collected and analyzed, but also that data on causal and contributing factors, mitigations, and associated performance is also considered. This more holistic view of performance supports analysis of realized events (a reactive approach) and opportunities to identify patterns and trends for proactive intervention. To accommodate both reactive and proactive action, performance measures should reflect both short-term and long-term performance. The insights gained from indicators support prioritization and focus on oversight actions as well as future data-driven decision-making (ICAO SM ICG, 2019a).

A commonly used approach to establish performance targets on which to draw insights from performance data and indicators is statistical analysis of trends. The specific type of statistics to be calculated is dependent on the underlying measured data and desired performance objectives. ICAO developed guidance on using statistical techniques to establish safety performance targets (also called action "triggers") and using the standard deviation principle to drive action (ICAO 9859, 2018). In this approach, historical data is used to determine mean performance, and performance targets or action triggers are defined by monthly variations, where trends beyond one or more standard deviations align with decisions for action.

In consideration of indicator performance to safety targets, care should be taken in the interpretation of data. This includes understanding that safety performance indicators, especially in complex systems like aviation, should not be treated as statistically independent.

Additional tools and techniques addressing performance measurement in safety oversight can be informed with related research in the context of Industry 4.0, or the fourth industrial revolution. This term is generally used in reference to the current and emerging era of advancements in analytics and automation, with significant increases in both connectivity and data. In developing and analyzing measures within Industry 4.0, the first step is to leverage a reference architecture that identifies all elements or actors within a system or domain, and for which different elements or levels within the reference architecture will be considered when developing and analyzing performance measures. For Industry 4.0, two standards that define this reference architecture include ISA-95 *Enterprise-Control System Integration* and ISO 22400 *Automation Systems and Integration – Key Performance Indicators for Manufacturing Operations Management* (Tambare, Meshram, Lee, Ramteke, & Imoize, 2018). In this approach, data and opportunities for performance measurement are assessed for each level and component within a system that captures or reports information, including real-time data processing components. ISO 22400 organizes measures into three broad categories.

- Key result indicators: lower-level and specific performance of individual components/actors
- Performance indicators: aggregates of key result indicators or a higher-level performance measure
- Key performance indicators: cross-process or actor results; this is reflective of overall process and performance outcomes

A tenet of Industry 4.0 is that data and data measures have an essential role in the development and management of quality. The additional creation, collection, and analysis of data is a key element of Industry 4.0 but also supports many elements of the corresponding Quality 4.0, where this references the desire to optimize and synthesize quality systems using modern technologies and techniques for quality improvements and quality culture (Tambare, Meshram, Lee, Ramteke, & Imoize, 2018).

6.3 Target level of safety

Another technique for establishing an acceptable level of performance for identified measures is a Target Level of Safety (TLS). A common method for establishing a TLS is to measure existing performance such as an accident rate at various severity levels and allocate risk to individual hazards (for example, dividing an accident rate by the number of hazards in the system). The result determines where acceptable risk is distinguished from unacceptable risk, frequently indicated in a risk matrix by a dividing line between red and yellow regions. Historical risk measurement may also serve as a baseline for future predictions and comparing the influence of any major proposed changes on safety risk. Furthermore, when data is used to quantify risk, statistical methods may be used to develop performance targets to validate the results of an initial risk analysis over time, and findings that invalidate an original analysis may indicate the need for oversight intervention.

7 Oversight effectiveness evaluation

Oversight effectiveness does not have a universal definition in the aviation safety community. ICAO SM ICG safety terminology refers to a level of safety as "a measurement of the effectiveness of a system's safety based on the probability of tolerable incidents that can occur" and "the degree of safety of a system" (ICAO SM ICG, 2022). While this definition does not directly mention safety oversight, the "system" within the definition can be extended to oversight, consistent with ICAO's characterization of a system as the "integrated set of constituent elements that are combined in an operational or support environment to accomplish a defined objective." ICAO also notes within the definition of safety performance that "effectiveness better describes the degree to which an activity of job is successful in producing the desired result," and in this context, safety oversight could be considered the "job."

Concepts for defining oversight effectiveness can also be related to broader definitions for measures of effectiveness. Within the Systems Engineering Body of Knowledge, a definition is given for Measures of Effectiveness (MOEs) as "the metrics by which an acquirer will measure satisfaction with products produced by the technical effort," consistent with IEEE's same definition (International Council on Systems Engineering, 2024). This definition implies that the customer has a bearing on whether a system such as safety oversight is effective and to what degree.

7.1 Independent audit

Independent audits can be used to gauge oversight effectiveness, with examples of this approach in aviation safety and many other regulatory domains. An independent role in assessing oversight effectiveness is important to capturing an objective perspective. ICAO established the Universal Safety Oversight Audit Program with the objective of helping states and ANSPs perform effective safety oversight in concert with ICAO Safety Management Manual (SMM) Document 9859 requirements which requires State Safety Programs to be assessed for effectiveness. This program focuses on critical elements of a safety oversight system where the audit framework includes the following phases (ICAO, n.d.), (Skybrary, n.d.):

- *Pre-audit phase:* administering an activity questionnaire and checklist to facilitate audit tailoring
- *On-site phase:* auditors visit with the safety oversight organization to assess the oversight system and capability, including organization, processes, procedures, and programs
- *Post-audit phase:* development and presentation of an audit report with corrective action plans to identify areas requiring improvement or attention

Focus areas for the USOAP audit include aircraft operations (OPS), airworthiness of aircraft (AIR), air navigation services (ANS), aerodromes, and ground aids (AGA), among others. ICAO auditors use an Effective implementation (EI) indicator as a measure of a regulator's safety oversight capability where a higher-scoring safety oversight system has a greater degree of compliance with ICAO provisions. Effective Implementation is evaluated relative to protocol questions intended to assess a state's safety oversight system, where the score is determined by the percent of satisfactory question responses (or data) compared to the total number of applicable protocol questions.   Though a quantitative Effective implementation score is produced, it is more a measure of compliance than effectiveness relative to achieving safety objectives or demonstrating customer (e.g., flying public) satisfaction (Gilbert, Lamont, & Bell, 2024).

7.2 Stakeholder surveys

The UK Department for Transport conducted a public survey in 2022 – 2023 to assess the effectiveness and efficiency of the UK Civil Aviation Authority's (CAA's) functions and structures, governance, and accountability (UK Department for Transport, 2023a). Stakeholders participating in the survey included UK CAA licensees (e.g., general aviation pilots), UK commercial airline passengers and cargo customers, airports' representatives, and airline and aviation industry participants regulated by the CAA, among others. Metadata on survey respondents was requested, including in what capacity the respondent was engaged in interactions with the CAA and whether those interactions took place in the past 5 years. A sample of questions in the UK Department for Transport's survey is presented in Table 6.

Table 6: Sample of UK CAA survey questions

• To what extent do you agree or disagree that the CAA has the right capability to fulfil its functions now and in the future?	• To what extent do you agree that the CAA engages effectively with its stakeholders (including customers)?

• To what extent do you agree or disagree that the CAA has the appropriate technical capability to make sound regulatory decisions? • To what extent do you agree or disagree that the CAA is appropriately structured to fulfill its functions / organized to fulfill its functions and why? • To what extent do you agree or disagree that the CAA is able to anticipate future regulatory requirements for the sector and prepare new regulatory frameworks in a timely fashion and why? • To what extent do you agree or disagree that the CAA effectively supports the aviation sector outside of its core regulatory functions (for example, by enabling economic growth, innovation, skills, and environmental sustainability in the aviation sector)?	• To what extent do you agree that the CAA seeks feedback to improve its performance and customer services? • How does the CAA's customer service compare to other aviation regulators? • To what extent do you agree or disagree that you are able to engage with CAA subject matter experts to gain timely advice on specific regulatory topics? • To what extent do you agree or disagree that the CAA effectively carries out its licensing duties? • To what extent do you agree or disagree that the CAA effectively regulates airlines with its current powers? • To what extent do you agree or disagree that the CAA effectively regulates airports and air traffic controls (economic regulation) with its current powers?

In July 2023, the UK Department for Transport published a report of findings and recommendations synthesized from survey responses on UK CAA functions and accountability (UK Department for Transport, 2023b). Report findings are organized into topics regarding the overall necessity of the CAA as an organization and the CAA's governance, accountability, efficiency, and efficacy. Some observations regarding the CAA's efficacy note public concerns regarding the CAA's ability to "effectively hold airlines to account" given its current powers. The report notes "a view from some consumer groups and some airlines that the behavior of a few is probably unlawful, and that the CAA relies on competition to address this rather than using its powers to at least ensure compliance with the law." Other findings concerning efficacy recommended that the CAA work with providers of new and emerging technology in developing changes to regulations "in an even more collaborative relationship than is the case in more traditional regulation" while being "clear as to the information it requires from its customers to make all regulatory decisions." Additionally, survey findings under the topic of UK CAA accountability noted consistent stakeholder feedback "that the CAA needs to increase its consumer focus" to which the CAA "acknowledged that it should be more systematic in discussing aviation consumer interests at board meetings." The recommendations for accountability go further, suggesting that a "consumer interest dashboard will strengthen governance in this area and give the opportunity to better utilize insight data to determine areas of focus that are important to aviation consumers."

7.3 Benchmarking

Benchmarking also serves as a technique for oversight effectiveness evaluation by helping to identify areas for improvement. Benchmarking uses a structured approach to identify and measure similarities and differences in the implementation of standards and practices among different organizations. In general, there are two types of benchmarking approaches:

- Standards benchmarking, which monitors the progress of an organization towards a defined set of best practice standards or policy targets
- Performance benchmarking, which compares performance across organizations considering a defined set of criteria or indicators

While some of the identified criteria can be considered to be important across organizations that implement SMSs or IMSs, others point to the need to apply a structured approach to benchmarking where the following are considered (Austrailian Government Productivity Commission, 2012):

- *Objectives*: this could be identifying types of functional activities, leading practices, etc.
- *Coverage*: scope addressed within a benchmarking activity
- *Measures / indicators*: definition of measures, which may be qualitative or quantitative. This should also consider data availability, ease of interpretation and comparability, robustness, and significance and relevance
- *Data management*: how data will be collected and applied
- *Reporting*: how benchmarking results will be reported

Additional information on benchmarking tools developed to reflect and evaluate the relative maturity of an oversight organization's capability was addressed in a previous literature review and is reported in the *Survey of Performance-Based and Risk-Based Oversight Strategies and Methods* (Gilbert, Lamont, & Bell, 2024).

8 Process modeling and automation

8.1 Model-based systems engineering

With the growing complexity in systems and data, more emphasis has been placed on the application of models to facilitate understanding and management of elements of a system, function, or service. As defined by the International Council on Systems Engineering (INCOSE), a model is a graphical, mathematical, or physical representation of a concept, service, system relationship, or structure. Generalization of concepts is applied so that models can represent functions or operations and facilitate understanding across a range of stakeholders. Leveraging models, model-based systems engineering (MBSE) is defined as "the formalized application of modeling to support system requirements, design, analysis, verification, and validation activities beginning in the conceptual design phase and continuing throughout development and later life cycle phases" (Hart, 2015). A number of tools and languages have been developed to facilitate the implementation of MBSE by supporting its architectural elements. For example, digital models include Mechanical Computer-Aided Design, Electronic Computer-Aided Design, Systems Modeling Language, and Unified Modeling Language are all intended to document and communicate the design and intended function of a system. Digital models and the tools that implement the models have helped to shift design and understanding away from a document-centric environment to model and data-centric environments. Principles of a model-based environment include:

- Sets of interconnected models that facilitate the capture of relationships and traceability
- Standardized languages, including notation, syntax, and semantics
- Shared information bases

To complement the architectural models, analytic models have been developed to quantify expected system performance and to provide insight into how systems might fail. As one example, Markov chains, which quantify the probability of transitioning between states, complement a SysML State Machine diagram. Similarly, the influence diagram and risk framework discussed in section 4.5 may be used to complement SysML activity diagrams.

A common pitfall of early implementations and application of MBSE was to create models without understanding the specific objectives the model addresses (Systems Planning & Analysis (SPA), 2019). To achieve the desired outcome, models need to leverage objectives and stakeholders to drive the format, granularity, and elements of models. Key capabilities of specific tools (generally commercial-off-the-shelf) to support MBSE include:

- Centralized requirements management and traceability
- Model-driven development including code development
- Simulation and testing
- Change management
- Collaboration and communication
- Reporting, analytics, and compliance support
- Customizable workflows and integration with other tools

There are a variety of tools and techniques with these MBSE capabilities as well as other functions such as IBM Rational Rhapsody, Siemens Teamcenter, Sparx Systems Enterprise Architect, Matlab Simulink, and SysML, among others (Visure, 2024).

Recent studies have demonstrated the potential for MBSE to support the assessment of safety oversight, for example, for demonstrating regulatory compliance. In (Bethel, 2021), MBSE methodology was demonstrated as an opportunity to move from document-based data exchange to model-based data exchange, where a model that stores and links airworthiness requirements to specific design and test data was demonstrated as a positive proof-of-concept. While there is a need for more comprehensive research in this area, these studies demonstrate the potential for MBSE products to provide supporting evidence for the evaluation of safety as well as modeling safety oversight requirements and processes to support the assessment of oversight functions relative to safety objectives.

8.2 Value chain analysis

Value Chain Analysis is a useful tool that has been commonly used to assess and improve a business service, product, or operation by identifying and thoroughly inspecting each step or function in the process from start to finish. Each step or function is viewed as a link in the chain, from resources and materials to the finalized product. Thorough analysis can be applied through the assignment of qualitative or quantitative value to identify and improve the weakest links and therefore improve the overall product, service, or operation. Within a business application, a value chain usually is organized into three overarching sectors:

- Inbound Logistics: Inbound logistics concerns the process involved in receiving the raw materials or inputs needed to develop the product. In the case of evaluation of a service, it can also address raw data collection.
- Operations: Operations is the main idea of a product chain and involves the steps for converting inputs into a product or service. It can include the development of software, algorithms, and procedures applied to transform raw data inputs.
- Outbound Logistics: This concerns the delivery of a service or packaging and distribution of the finalized product to the customer.

An example is shown in IATA's airline value chain reference model from 2020 which addresses airline operations, management, and support functions as well as airline customer touch points. The value chain is developed in Enterprise Architect and backed by Unified Modeling Language (UML) schema that captures definitions of each function along with traceability to other parts of the value chain and interrelated conceptual and logical models (IATA, 2020). With this foundation, changes (e.g., improvements, modifications, disruptions) to one or more functions in the chain can be readily cross-referenced with dependent functions to facilitate change impact analysis.

A significant oversight challenge noted in FAA AVS Future of Oversight Steering Committee discussions in 2024 entails the gap between the flying public's expectations for aviation safety and the regulations and policies that govern safety oversight. When the flying public experiences traumatic incidents on an aircraft (especially when echoed by intense media coverage), their expectations and perceived personal risk exposure do not always align with the classification and treatment of such events for the purpose of tracking aviation safety occurrences and guiding oversight focus. Value chain analysis and other process modeling techniques may offer an approach to capture the flying public's expectations and risk tolerances (as characterized by utility theory, for example) for the safety of flight which can be represented in a similar fashion to the customer touch points in IATA's airline reference model. The value chain can be used to connect the execution and delivery of safety oversight services to those customer touch points with the opportunity to model as-is and to-be versions of the oversight system.

8.3 Business process modeling

Business Process Modeling (BPM) is a management philosophy, supported by a range of methods, techniques, and tools, for assessing and improving the effectiveness of an organization. It can be used in multiple ways to evaluate and identify opportunities to tune and evolve an organization to improve efficiency and effectiveness through structured definition and decomposition of business processes (Reijers, 2021). Foundational elements of BPM to support these objectives are visualizing processes and analyzing performance. BPM can be used to create visual representations, such as flow charts and diagrams, of an organization's processes or process flow. The visualizations help to understand sequences of activities, decision points, and interactions both within an organization and between an organization and external participants. Since BPM process models can be used to identify how participants in a process interact, and for what purpose, they can be used to promote better communication and collaboration between participants. Performance metrics or indicators can also be defined for each process to measure process effectiveness relative to quality, efficiency, and/or safety. Performance indicator data may also be used to support the measurement of performance to organizational goals, which can include safety-oriented goals. Key features of software tools that support business process modeling include:

- Capture and visualization of business process activities and relationships
- Facilitation of data sharing and collaboration to collect feedback and drive process updates

- Analytics for the performance of individual processes or process flows; this supports a data-driven approach to improving or redesigning processes to improve effectiveness, including addressing risk and meeting safety objectives.

While defining and assigning performance indicators to individual processes and process flows is relatively straightforward, it can be challenging to consider business process performance in the context of addressing risk. More recent research on process modeling has focused on the application of BPM for risk modeling (e.g. Risk-aware BPM (R-BPM)) (Lamine, et al., 2020). While the literature focus is generally on *business risk* management and risk lifecycle within an organization, there are elements that can be extended to modeling *safety risk management* within and across organizations (including elements of safety oversight organizations) and to understanding organizational elements that impact the risk management system. Withing R-BPM and similar concepts, efforts to evaluate business processes are overlaid with risk process activities, capturing relationships between organizational roles and business processes with the larger context of risk management activities of identifying, analyzing, treating, and monitoring of risk. To do this, the risk management perspective needs to consider activities within the whole scope of the analyzed environment. For example, for the US National Airspace System, risk management needs to be considered from the holistic perspective that accounts for all systems, persons, operations, environments, and organizations that have a role in the NAS. An understanding of the relationship between a specific organizational role or process (e.g., a safety oversight function) and its overall responsibility in the risk management process can be used as another input for the evaluation of business process performance (i.e., considering the performance of a process to its risk management responsibilities)).

BPM tool outputs can also benefit process standardization within oversight organizations and comparison of performance across organizations relative to an accepted standard. Creating standardized processes shared across oversight organizations, especially when oversight actions are common to a regulated entity or product, drives consistency and reliability which are important for maintaining quality and achieving desired outcomes. Benchmarking tools, described in section 7.3 of this report, can also be used to compare process performance to an accepted standard or to other organizations as part of BPM.

BPM tools can be used to explore newly proposed or modified organizational processes and their impact on an oversight organization. An organization must first define new or changed objectives. Next, detailed models of current safety oversight processes using flowcharts, swim lane diagrams, or other BPM tools are created (if they do not already exist). If a new oversight process is already defined, this too can be modeled and used as a foundation for identifying organizational or process risk.

Process models are analyzed to determine if and how they contribute to organizational objectives, and to identify gaps, redundancies, and inefficiencies. Of note is that an organization should be mindful that when considering processes that may appear redundant, opportunities and insight gained from multiple, independent assessments of a common problem space can be useful and support confidence in meeting performance objectives. Processes can be tuned or redefined to address the identified gaps or inefficiencies. Finally, when changes are made, BPM tools include monitoring with performance indicators that can measure organizational outcomes and performance, supporting validation of changes and continuous improvement.

Throughout the BPM process, there are BPM-specific software tools that can be used to collect, manage, and report process data; these tools can be used independently or possibly integrated with other management system tools. Some examples include Appian BPM Suite, Nintex workflow automation, IBM Business Automation Workflow, and CMW Platform. While specific software tools offer different solutions to support the capture, storage, and visualization of process information and supporting performance data, the power in achieving meaningful outcomes is applying the *techniques* of BPM, leveraging clear and complete data regardless of how the data is managed and visualized.

8.4 Robotic process automation

As business processes are evaluated for the opportunity to improve efficiency, quality, and safety, there is also a case for evaluating the digitization and automation of repetitive organizational tasks. Robotic Process Automation (RPA) can reduce workload related to data cleaning and transformation in preparation for data analysis; non-digital data transfer and transformation; and repetitive and administrative workflow processes to allow the workforce to focus on tasks that are more complex and cognitively challenging. RPA uses software tasks that are either running unattended (e.g., at all times), with reliance on a defined data structure to perform specific workflow activities or attended meaning that they are enabled when directed by a person (Farinha, Pereira, & Almedia, 2023). Generally, there are three components within the RPA environment, as shown in Figure 12 below.

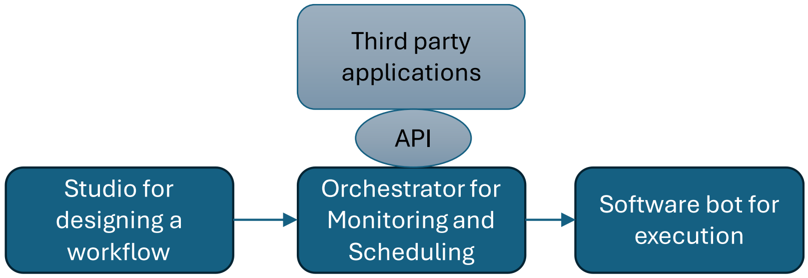


Figure 12. Components in an RPA environment

A summary of component functions includes (Rizvi & Srivastava, 2023):

- Studio for design: This includes the capability to define a workflow to be executed, including process steps, interfaces, and data
- Orchestrator: This monitors and schedules task activities when needed and provides software bot management interfaces.
- Software bot: These are programmed software scripts that are launched and run to fulfill the conduct of organizational activities entailing tasks such as:
    - Web scraping: Gathering content from a variety of sources and aggregating it
    - Data collection/transformation: This can include copying data from one place to another, cleaning and transforming data, and processing voice into text
    - Authentication or compliance: This can be used to validate data and to evaluate data relative to standards or performance targets
    - System integration: Merging similar data from disparate sources.

There are a range of commercial products that facilitate the design, implementation, and operation of robotic process automation and can interface to elements within the RPA environment via application programming interfaces (i.e., APIs); examples include UiPath (for discovery and automation of workflows), Automation Anywhere, Hyland RPA to facilitate reporting, and Microsoft Power Automate.

A challenge can be identifying which organizational processes are suitable candidates for the application of robotic process automation. Researchers have developed a checklist-based tool that defines criteria to consider when assessing RPA for specific business processes. These criteria and associated references are provided in Table 7 (Farinha, Pereira, & Almedia, 2023).

Table 7. Checklist Tool Criteria to Assess Processes for Business Process Automation

Category	Criterion	Description
Data	Input and output data	The data structure used as inputs and outputs should be standardized and semi-structured

Category	Criterion	Description
	Data security	Data access and manipulation already comply with security best practices.
	Data digitalization	If the data is being digitally managed
	Test data	If there is any data to be used to test before deployment
	Number of applications involved	Number of applications used by the process
Environment	Applications similarity	If the involved applications have a similar structure or differ a lot. For instance, similar APIs or programming languages
	Applications stability	Involved applications are stable, and fewer updates are expected throughout the time
	Applications requirements	The complexity of requirements to connect with applications
Human resources	Human effort	How much human effort is needed to complete a transaction before automatization
	Human involvement	How much human involvement is needed to complete a transaction after automatization
	Number of users	How many users are involved in the process may represent significant savings in resource relocation
	Cognitive requirements	How many human interventions/decisions are required to complete a transaction in a process
	Human error	The amount of human error that a business process has
Governance	Process cost	How much is the cost of the process
	Savings	Any benefit that can come with the automatization
	SLA impact	If the automation will increase the SLA compliance rate
	Time-consuming	How much time is required to finish a transaction
	Accurate process description	If all the process documentation exists and is clear
	Number of robots allowed	Each enterprise contract with the RPA software provided may limit the number of possible robots in parallel
	Reusability	If it is possible to reuse part of other RPA developments, or if the current development may be easily reused
	Efficiency	If it is expected to increase the productivity and quality of the process
	Feasibility	How complex is the automatization given the overall process complexity?
Structure	Number of process steps	If the process has a considerable number of steps or not
	Process complexity	How complex the process is considering its entire context like human involvement, applications involved, and process steps, among others.
	Number of exceptions	Number of possible exceptions that a process can have
	Process stability	If the process did not suffer any significant change in the past 12–18 months, indicate that it is stable and fewer updates are expected throughout the time

Category	Criterion	Description
	Process standardized	The process should already be standardized; otherwise, the development will take a lot longer, and the robot will face a bigger number of exceptions that were not mentioned while in the development
	Repetitive	Usually already standardized, but the process always follows the same ruled-based workflow
	Number of transactions	Processes that originate high amounts of transactions (number of times the bot does a task) are good candidates for automation
	Predictability of outcomes	Despite some known outputs being structured, others may be unpredictable
	Rule-based process	The process is dominated by business rules which are already contemplated in the process
	Automation type	Attended processes run in user machines and are more dependent on human intervention. Unattended processes run in virtual machines and work 24/7

An area of oversight where Robotic Process Automation has been explored quite extensively is auditing. In this area, there is evidence in both experiments and pilot programs that the use of RPA demonstrates efficiency gains and accuracy improvements (Solanki, Mehta, & Shukla, 2024). For example, manual audit processes that are repetitive and standardized such as audit data collection, storage, and reporting have the potential to reduce oversight workload through automation. Additionally, as RPA leverages the digitization of data and processes, its implementation facilitates integration with artificial intelligence and machine learning, providing additional opportunities for drawing insights within oversight audits.

Within the aviation industry, applications for RPA are also emerging in aircraft maintenance programs. RPA can be used to identify differences between maintenance records and an approved maintenance program, translate maintenance program changes into software, produce routine maintenance work orders, and initiate reliability reports among other functions (Barrera, 2022).

9 Safety culture and human performance assessment

9.1 Safety culture assessment

There is a growing recognition of the importance of safety culture within safety organizations, especially in aviation where human behaviors and organizational values influence safety:

> "In aviation, there is a core underlying tenet that 'people create safety' and keep the skies and passengers safe, based on a robust industry-wide safety culture." (Kirwan, 2024)

Across industries, there are varying definitions of safety culture, and as a result, implementation approaches and concepts for measurement can vary considerably. Before establishing specific assessment approaches and tools, a clear definition of safety culture is needed. ICAO SM ICG defines safety culture as “an enduring set of values, norms, attitudes, and practices within an organization concerned with minimizing exposure of the workforce and the general public to dangerous or hazardous conditions” (ICAO SM ICG, 2022). ICAO’s definition is similar to an earlier definition proposed by the Civil Air Navigation Services Organization (CANSO) in which “safety culture refers to the enduring value, priority and commitment placed on safety by every individual and every group at every level of the organization. Safety culture reflects the individual, group, and organizational attitudes, norms, and behaviors related to the safe provision of air navigation services” (CANSO, 2008).

Approaches for measuring safety culture in aviation are built on work in the nuclear power and oil and gas industries. Following accidents in Europe in 2001 and 2002, joint efforts between EUROCONTROL and several universities resulted in a safety culture measurement questionnaire, and after pilot testing with targeted air traffic providers, rolled out across Europe in 2005 (Kirwan, 2024). The questionnaire focused on safety culture “dimensions,” including:

- Management commitment to safety
- Collaboration and involvement
- Just culture and reporting
- Communication and learning
- Colleague commitment to safety
- Risk handling
- Staff and equipment
- Procedures and training

In recent years, more focus has been applied to examining methods and tools that assess safety culture. This is true across organizations that implement or evaluate aviation safety management systems, including the FAA. For example, the FAA’s Maintenance Safety Culture Assessment and Improvement Toolkit (M-SCAIT) provides a handbook (released in May 2023) that identifies methods for assessing culture, including surveys and interviews, and provides guidance on measurements related to Job Resources, Job Demands, Employee Satisfaction, Employee Well-Being, and Organizational Outcomes (Key K. , Hu, Schroeder, & Choi, 2023).

Also published in May 2023, the FAA NextGen Human Factors Division sponsored “Safety Culture Assessment and Continuous Improvement in Aviation: A Literature Review.” This study explores literature addressing measurement of culture but also barriers for intervention and improvement noting that “if culture is not responsive to interventions there is no point in measuring it” (Key K. , Hu, Choi, & Schroeder, 2023). Types of safety culture measurement tools addressed in this study are summarized in Table 8.

Table 8. Safety culture measurement tools with strengths and weaknesses reproduced from (Key K. , Hu, Choi, & Schroeder, 2023)

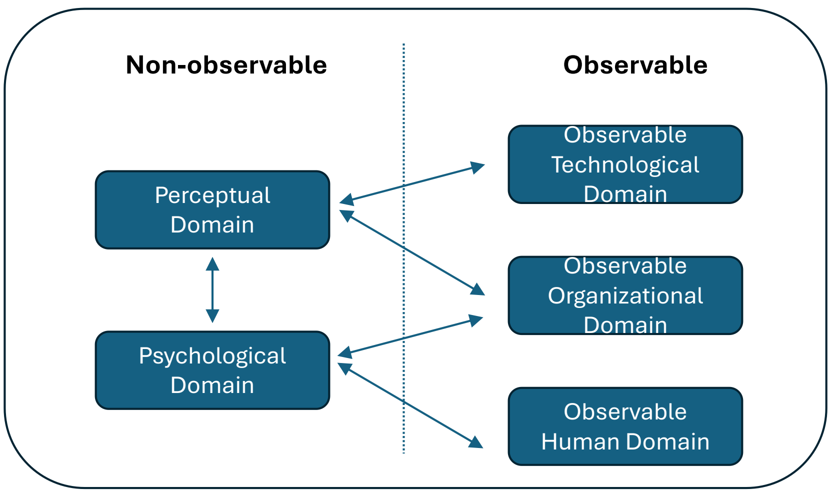
Tool	Strengths	Weaknesses
Survey	Employees identify their areas of concern; promotes “ownership”; can be administered quickly and supports aggregated view of attitudes and perceptions	Results are based on self-reported attitudes and perceptions
Interview / Focus Groups	Provides detailed information as a result of active discussion; employees can identify areas of concern	Attitudes or perceptions may be influenced by unrelated factors; more time-consuming than questionnaires; participants may not reflect the general view of the larger team or organization; structed process and content is needed
Observations	Supports interaction and cooperation between observers and employees. If observers are peers or trained persons within an organization, they may be able to identify information not as accessible to management or outsiders	Knowledgeable and trained observers are needed; observation may be intrusive and impact normal tasks and operations.
Incident/accident reviews	Leverages factual evidence	Relies on the competency of those performing the root cause analysis; because events are rare, sparse data may not be representative of a population

While general strengths and weaknesses are reported above, any tool that is not proficiently and adeptly executed, even with the best intention, will not provide the desired insights into safety culture. The previous literature survey also examines a mixed-tool approach, with the objective of obtaining a representative sample of target employees (Key K. , Hu, Choi, & Schroeder, 2023). Also of note from this previous body of work is the need to tailor safety culture assessment to address organization-specific roles, topics, and safety responsibility areas, among other considerations. Several methods can be used to establish confidence in the validity of survey, interview, and observation questions, including the use of a construct validity table and subject matter experts to review statements or questions for adequacy in measuring applicable safety culture concepts (Hammond, King, Joe, & Miller, 2023).

A trend in safety culture assessment is to apply more holistic approaches, considering multiple elements that comprise or influence culture, including leadership commitment, employee perceptions, reporting behaviors, and organizational policies (Reniers, Ponnet, & Van Nunen, 2022). In a holistic approach, the direct collection of data from across an organization including different roles and responsibilities provides a clearer understanding of safety culture perceptions. This is of particular importance to assess both senior and middle management, as their actions and decisions have been shown to have a dramatic ripple effect on safety culture within an organization (Kirwan, 2024).

An example tool for holistic culture assessment is integrated safety culture assessment, which collectively addresses all elements of interest organized into five primary domains with both observable and non-observable culture elements. Figure 13 provides an overview of this type of holistic approach, adapted from "Measuring Safety Culture Using an Integrative Approach: The Development of a Comprehensive Conceptual Framework and an Applied Safety Culture Assessment Instrument" (Reniers, Ponnet, & Van Nunen, 2022).

Figure 13: Elements of an Integrated Safety Culture Assessment (ISCA)



Within the non-observable behaviors, data is collected on users' perceptions on items such as handling anomalous conditions; supervisory and management leadership and commitment; employee commitment; safety department commitment and impact; and supporting environment (including time, people, training, and safety rules). Elements in the psychological domain address elements including trust, intention, overall safety knowledge and competency, and personal priorities. This sample framework (Reniers, Ponnet, & Van Nunen, 2022) organizes observable behavior considerations. Common means for data collection for non-observable behaviors is through surveys or interviews. When developing survey and interview questions, it is important to leverage experts in the field and allow time for testing on small trial groups to ensure that the desired outcomes and data are achieved. Factors that impact final questions are determining the number of positive and negatively worked statements; ensuring the questions account for the needs and context of the organization under assessment; and tailoring for specific roles in the organization that the interviewee may perform.

Techniques to collect data for observable behaviors include inspection, monitoring, and data review where observable behaviors are categorized into three domains, including:

- *Technological domain*: the physical working environment and interactions with systems and environment.
- *Organizational domain*: safety policy and goals, safety performance, safety communication and transparency, safety inspection, and safety procedures.
- *Human domain*: safety compliance and safety participation

Another focus of more recent research is exploring the correlation between safety culture measurement safety outcomes and leading safety indicators. This requires investment in a long-term vision of safety culture measurement and identifying the impact of safety culture initiatives on safety outcomes and indicators. Dependencies among safety culture traits were explored in one study and then used to identify key performance measures reflective of safety outcomes in U.S. nuclear waste remediation (Hammond, King, Joe, & Miller, 2023). Collected data was used to evaluate the correlation of performance measures reported monthly with safety culture survey factors. The analysis found that certain measures of overall organizational performance such as evaluation and resolution of safety issues were predictive of improved safety culture; however, certain management initiatives and techniques intended to improve safety culture sometimes led to negative and unintended safety performance outcomes. While such analyses have challenges in obtaining complete and consistent data (especially for low frequency-of-occurrence performance indicators), there are opportunities to identify safety culture factor relationships with performance indicators and to quantitatively measure the impact of safety culture initiatives on outcomes and leading indicators.

9.2 Human performance and interaction assessment

Human performance is the measure of what people can achieve specific to their role within a system or organizational environment. There is a wide range of traditional tools and techniques leveraged for human performance measurement, generally selected and tailored for a specific task or measurement of interest. Examples of such tools and techniques for human performance measurement are reported in the US Department of Energy's Human Performance Improvement Handbook as shown in Figure 14 (US Department of Energy, 2009).

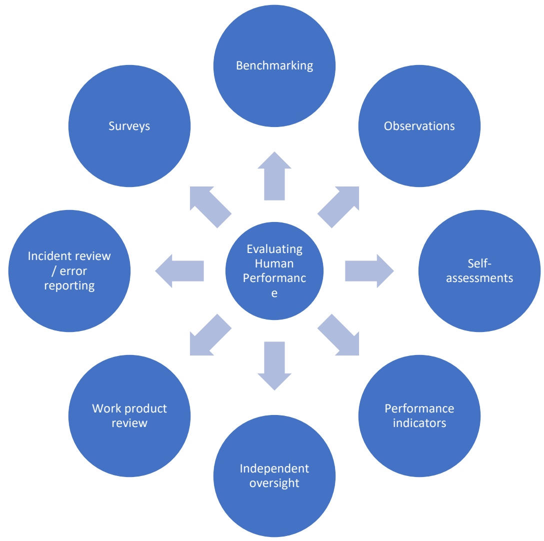


Figure 14. Tools supporting human performance measurement

One tool that supports human performance analysis but is not mentioned above is barrier models, where human barriers to be considered include personnel attributes, resources for task performance, physical work environment, and organizational work environment (Barnes & Haagensen, 2002). These can be associated with human-performance-specific threats and evaluated when an issue or off-nominal condition occurs. More on barrier modeling is addressed in Section 4.3.

As the tools in Figure 14 are generally well understood and widely applied, this report focuses instead on emerging concepts and tools that may be used for or inform human performance measurement. One such area considers frameworks and tools to support comparative assessment of machine learning algorithms with human performance. While the aim of these tools may not be to assess human performance, the research approaches explored to compare machine learning algorithms with human performance have advanced concepts in human performance evaluation. A framework identifying guiding principles for human performance measurement is put forward in Cowley, et al. (2022). This framework leverages component findings of previous, similar research and includes:

- Study of specific stimuli and responses in pilot studies
- Recruiting and use of large subject pools
- Applying measurement control for specific performance strategies or attributes (e.g., memorization, language ability, attention)
- Creating benchmarks to evaluate performance models and measures
- Conducting research outside of a laboratory (using real or nearly real operating environments)

- Collecting data on measurement constraints (including as perceived by participants)
- Reporting analysis on measurements of interest but on other variations as well as patterns of false positives and negatives
- Collecting demographic and task-related subjective data (e.g., with tools such as NASA Task Load Index)

While these principles themselves do not constitute a human performance measurement tool, they do provide criteria that can be used to assess measurement tools.

Another area of emerging research focuses on the collection and application of objective data for human performance measurement. A recent maritime study (Fan & Yang, 2023) investigated opportunities to supplement subjective assessment of human performance with tools and techniques that provide *objective* data. This study uncovered biases of trainees in performing actions they perceived would be expected by the assessor, impacting the way the assessor retrieved and perceived data. This drove the implementation of an approach where computer-based assessments were integrated into assessment programs to produce a more objective measurement. The maritime human performance measurement study also reported a positive correlation between individual factors (e.g., mental workload, emotion, fatigue, time pressure) and work behaviors and decision-making, where there is a desire to quantify potential individual risk factors.

Additional emerging tools and techniques supporting the generation of objective data for human performance measurement are found in the field of human reliability analysis. Here, the research focus is on investigating the use of sensors and connected devices to collect data from humans performing tasks and to understand and objectively quantify human behavior and human/system interfaces. Data captured from these measures provides an opportunity to create or inform benchmarks that can be used to assess other persons performing in similar roles or environments or to identify and assess behavioral or human performance variations. Example tools and techniques that use sensors and connection devices for human performance measurement include (Fan & Yang, 2023):

- Electrocardiography: a study of electrical activity associated with the human heart.
- Electromyography: study of electrical activity associated with nerves and muscles
- Electroencephalography: study of electrical activity within the human brain
- Skin electric response: study of changes in electrical properties in the skin
- Eye-tracking: observation and measurement associated with eye movements, pupil dilation, blinking, and point of focus

Supplementing these are new psychophysiological methods aiming for additional objective and data-driven assessments. For example, functional Near-Infrared Spectroscopy uses a wearable sensor to measure cortical activity when a user performs real-world tasks and scenarios where outputs can be used directly or input to artificial neural network models used to predict and classify trainee levels of qualification (Fan & Yang, 2023). The study found strong evidence suggesting valid prediction of the experience of evaluated participants based on collected data (Fan & Yang, 2023). While additional research is needed to independently validate these types of results, it does point to the potential of applying techniques for objective measurement techniques in combination with data models to measure and predict human performance.

Another focus of human performance research is on the evaluation of inter-organizational collaboration, both within an oversight organization and between a regulator and regulated entity. An area of interest is the organic evolution of collaboration vs. directed or engineered collaborative relationships; here there is the opportunity to promote a *virtuous cycle of collaboration* from directed action and information exchanges at appropriate frequencies with stakeholder acceptance (Vestola & Eriksson, 2023). A virtuous cycle refers to chains of events that reinforce themselves through a continuous feedback loop, resulting in favorable outcomes. Organizations that discover the key elements of their unique virtuous cycle find themselves in an advantageous feedback loop that drives success.

Some techniques that can be used to engineer collaboration and build virtuous cycle relationships include start-up/collaboration workshops, joint objectives, team building, co-location in joint project offices, and partnering facilitators (Vestola & Eriksson, 2023). While these can be valuable in working to share virtuous cycle relationships, informal partnering and informal collaboration naturally emerge over time, especially with mutual trust, shared values, openness in communication, and leveraging of previous experiences. These elements can be used to help assess existing relationships in combination with engineering relationship factors (e.g., implementation of workshops, joint objectives, and team building).

A checklist-based approach for human performance measurement is provided in the US Nuclear Regulatory Commission's report on "The Human Performance Evaluation Process: A resource for reviewing the identification and resolution of human performance problems." The report provides an evaluation checklist to identify and assess human performance issues and off-nominal performance. An underlying assumption in the application of the checklist is that a wide consideration of participants should be considered when performing an assessment as initially, it may be unclear that the actions of a specific person have any role in an identified issue or off-nominal performance. Additionally, an open and just culture environment is needed to ensure that participants in such an evaluation feel comfortable sharing clear and complete answers to questions. The checklists address four primary topics, using open-ended questions to identify and describe off-nominal performance conditions, including problem identification and characterization, incident investigation, causal analysis, and corrective actions. When considering issues or off-nominal performance, it is noted that the following topics should be considered: (1) errors of commission, in which incorrect actions were taken, (2) errors of omission, in which required actions were not taken, (3) extraneous acts, in which an action that was not required was taken, and (4) missed error-recovery opportunities, comprised of actions which could have corrected previous errors (Barnes & Haagensen, 2002).

10 Summary and conclusions

To address FAA Future of Oversight research objectives, this report presents a compendium of safety oversight tools, techniques, and technologies used by international aviation authorities and the aviation industry, among other techniques relevant to aviation safety oversight processes. Tools selected for the survey are based on guidance and findings from AVS interviews and AVS Future of Oversight Steering Committee discussions as well as applicability to the SMS pillars for safety policy, risk management, assurance, and promotion as defined by ICAO. The tools surveyed, including techniques and technologies, are organized by functions relevant to safety oversight including:

- SMS evaluation
- Safety analytical tools and techniques
- Safety information management
- Safety objectives and safety performance measurement
- Oversight effectiveness evaluation
- Process modeling and automation
- Safety culture and human performance assessment

Key observations, including some recommendations for further research, are as follows:

SMS evaluation

- SMS evaluation tools intended to assist oversight authorities with effectiveness evaluation focus more on implementation, understanding, and commitment to required elements of SMS. There is an opportunity to complement this approach with effectiveness evaluation techniques that examine the achievement of safety objectives and corresponding safety performance indicators; maturity of processes that support SMS; and capabilities to self-identify issues to be addressed and opportunities for improvement.

Safety analytical tools and techniques

- Artificial intelligence and machine learning (AI / ML) applications require data science expertise and organizational investment to curate and maintain data stores and guide iterative improvements to managed models. Oversight organization-specific needs, benefits, and readiness for AI / ML adoption should be examined with cross-cutting needs captured in requirements that can be evaluated for centralized vs. distributed AI / ML infrastructure implementation.
- Artificial intelligence and machine learning techniques, coupled with big data amassed from aviation safety events and operations, offer insights into emerging and unknown areas of risk. Strategic planning is needed to ensure that agreements (or regulations) are in place with the industry to enable oversight data collection for new entrant technologies and operations.
- While precursor and root cause analysis can be conducted for individual aviation safety incidents and undesired safety events, there is a broader benefit to aviation safety when such data is examined collectively across regulated entities and operations for systemic patterns. Analytical tools, including AI / ML, should be explored for use in precursor and root cause analysis with actual oversight data to determine performance and benefits vs. investment in infrastructure and expertise as well as recommendations for additional oversight data collection.
- Data on identified safety precursors and root causes should be combined with information on planned changes to aviation systems and operations. A predictive capability is needed to identify changes with greater risk to aviation safety to offer decision support in prioritizing such changes for proactive oversight.

Safety information management

- Blockchain technology can be leveraged to securely develop and maintain certification bases and records of changes, repairs, and other transactions over an aircraft, product, or system life cycle. Further research is needed to explore the feasibility and benefits of blockchain to allow multiple oversight organizations and the regulated entity to interact with a common and comprehensive record of certification data.
- Aviation safety oversight effectiveness and efficiency may benefit from establishing an oversight data exchange model to organize, standardize, and facilitate information

sharing across regulatory and industry organizations. Such an exchange model could incorporate the FAA’s Aerospace Hazard Classification System for management of aviation safety risk data.

- FAA's progress in establishing and socializing the Aerospace Hazard Classification System should be capitalized on by further expansion within AVS and the broader FAA community with consideration for recommending adoption by organizations that do not have an established aviation taxonomy.
- Tools to facilitate the digital exchange of data between oversight organizations and regulated entities or across individual oversight functions may enable more comprehensive and efficient data sharing and provide a foundation for a real-time understanding of aviation safety risk and decision analysis for oversight actions.

Safety objectives and safety performance measurement

- The use of risk-based approaches as well as process modeling can inform the identification of safety objectives for oversight and connect those objectives to aviation safety risks and specific oversight activities and functions. In turn, this connectivity can promote safety oversight workforce understanding and adoption of performance-based oversight and how individual oversight work and activities contribute to overall aviation safety.
- Identification of high-level, goal-oriented safety objectives can be translated into both leading safety indicators and outcome-focused performance measures. Research to define safety performance indicators and measures holistically for FAA aviation safety oversight is needed along with a framework to allocate or trace higher-level objectives, indicators, and measures to lower-level supporting oversight activities.
- The collective set of safety performance indicators for safety oversight should address all elements of SMS, providing a holistic view of safety performance.

Oversight effectiveness evaluation

- Approaches for independent audit of oversight effectiveness focus on the degree of implementation according to various maturity scales. To complement this approach, research into performance-based evaluation of oversight effectiveness is needed. Risk-based models, including barrier analysis, offer an opportunity to examine how oversight decisions and actions can be applied as risk mitigations. Further research is needed to determine if such models can provide a framework for not only evaluating individual oversight decisions and actions but also overall oversight effectiveness in reducing aviation safety risk.

Process modeling and automation

- Value chain and other process modeling and analysis techniques could be explored to address the gap between the flying public's expectations for aviation safety and current safety oversight functions. Process modeling techniques offer an approach to capture the flying public's risk tolerances in comparison with the conduct and delivery of oversight services modeled in as-is and to-be versions of the oversight system.
- An increased volume and digitization of aviation data for safety oversight analysis entails an increased workload on the FAA's oversight workforce. To offset part of this workload, repetitive manual data processes could be examined for suitability to automate using techniques such as Robotic Process Automation.
- Research is needed to produce oversight guidance on how to analyze regulated entity manufacturing, quality, safety management, and change management processes for safety risks. As part of FAA oversight, an independent safety risk analysis of entity management processes can be conducted to ensure that processes are robust to errors of omission, commission, timing, and abnormal and non-nominal conditions such as production surges. Formal tracking and application of lessons learned from prior process lapses across all regulated entities can enhance safety risk analysis of new and modified management processes.
- Collective oversight data sets across regulated entities also provides a foundation to validate accepted means of compliance and to develop shared best practice principles that can support for a center of excellence concept.

Safety culture and human performance assessment

- Further research is needed to investigate the potential identification of safety objectives and safety performance indicators for safety culture evaluation and to validate whether safety performance measurement can be used to predict and improve safety culture. This research should consider the evaluation of safety culture throughout all levels of an organization and the full supply chain where applicable.
- Precursors and root causes for historical aviation safety incidents and undesired safety events that reflect safety culture deficiencies could be used to inform objectives and performance indicators for future safety culture evaluation.

11 References

14 CFR 5.5. (2023, January 25). *Definitions*. Retrieved from law.cornell.edu: https://www.law.cornell.edu/cfr/text/14/5.5

Ackley, J. e. (2020). Supervised Learning Approach for Safety Event Precursor Identification in Commercial Aviation. *AIAA: Using Big Data to Increase Safety.*

Aerotime Advisory Team. (2023, August 3). *Big data in aviation management: revolutionizing decision-making*. Retrieved from Aerotime: https://www.aerotime.aero/articles/big-data-in-aviation-management-revolutionizing-decision-making

Ahmadisheykhsarmast, S., Aminbakhsh, S., Sonmez, R., & Uysal, F. (2023). A transformative solution for construction safety: Blockchain-based system for accident information management. *Journal of Instrustrial Information, 35*. doi:https://doi.org/10.1016/j.jii.2023.100491.

Air France Industries. (2024). *SkyThread: AFI KLM E&M and Parker Aerospace Group reach major milestone in deploying SkyThread's Blockchain-Based aircraft Parts Track and Trace platform for 787 fleet.* Retrieved from https://www.afiklmem.com/en/press-release/25072024_afiklmem-and-parker-aerospace-group-reach-major-milestone-in-deploying-skythread-s-blockchain-based-aircraft-parts-track-and-trace-platform-for-787-fleet

Austrailian Government Productivity Commission. (2012). *Performance Benchmarking of Austrailian Business Regulation: The Role of Local Government as Regulator.* Melbourne: Commonweath of Austrailia.

Australia Civil Aviation Safety Authority. (n.d.). *Using our bowtie risk analysis*. Retrieved from CASA: https://www.casa.gov.au/search?keys=bowtie

Australian Government. (2021-2023). Australian National Aviation Safety Plan. Retrieved from https://www.infrastructure.gov.au/department/media/publications/australian-national-aviation-safety-plan-2021-2023

Australian Transport Safety Bureau. (2020). *ATSB Transport Safety Report Aviation Occurrence Investigation AO-2019-035, 'Tail rotor pedal failure and collision with terrain involving Robinson R22 helicopter, VH-HHQ near Kutchera Station, Queensland, on 22 June 2019.* Queensland: ATSB. Retrieved from https://www.atsb.gov.au/media/5778587/ao-2019-035_final.pdf

Barnes, v., & Haagensen, B. (2002). *The Human Performance Evaluation Process: A Resource for Reviewing the Identification and Resolution of Human Performance Problems.* Office of Nuclear Regulatory Research, Division of Engineering Technology. Washington, DC: U.S. Nuclear Regulatory Commission. Retrieved from https://www.nrc.gov/docs/ML0209/ML020930054.pdf

Barrera, D. (2022). *Aircraft Maintenance Programs.* Springer.

Bernard, B. (2024). Responsive Regulation, Trust, and Intrinsic Motivation Within the Nuclear Industry: Impacts of a Safety Culture Tool. In J.-C. Le Coze, & B. Journé, *The Regulator–Regulatee Relationship in High-Hazard Industry Sectors: New Actors and New Viewpoints in a Conservative Landscape* (p. Chapter 11). Nantes, France: Foundation for an Industrial Safety Culture. doi:https://doi.org/10.1007/978-3-031-49570-0

Bethel, B. M. (2021). *Model Based Systems Engineering Demonstrator for Air Worthiness with Performance Data of Department of Defense Rotorcraft.* Air Force Institute of Technology. Retrieved from https://scholar.afit.edu/etd/4939/

Britten, T. (2022, December 14). *How to Create Safety Objectives in Aviation SMS with Examples*. Retrieved from SMSPro: https://aviationsafetyblog.asms-pro.com/blog/how-to-create-safety-objectives-in-aviation-sms-programs

Buselli, I., Oneto, L., Dambra, C., Gallego, C. V., Martinez, M. G., Smoker, A., . . . Martino, P. R. (2022, February). Natural language processing for aviation safety: extracting knowledge from publicly-available loss of separation reports. *National Library of Medicine*. doi:10.12688/openreseurope.14040.2

CANSO. (2008). *Safety Culture Definition and Enhancement Process*.

Coglianese, C. (2024). Rule Design: Defining the Regulator-Regulatee Relationship. In J.-C. Le Coze, & B. Journé, *The Regulator-Regulatee Relationship in High-Hazard Industry Sectors: New Actors and New Viewpoints in a Conservative Landscape* (p. Chapter 10). Cham, Switzerland: Foundation for an Industrial Safety Culture. doi:https://doi.org/10.1007/978-3-031-49570-0

Cowley, H. P., Natter, M., Gray-Roncal, K., Rhodes, R., Johnson, E., Drenkow, N., . . . Gray-Roncal, W. (2022, March 31). A framework for rigorous evaluation of human performance in human and machine learning comparison studies. *Scientific Reports, 12*(5444). doi:https://doi.org/10.1038/s41598-022-08078-3

Das, S., Matthews, B., & Lawrence, R. (2011). Fleet Level Anomaly Detection of Aviation Safety Data. *IEEE Conference on Prognostics and Health Management* (pp. 1-10). IEEE. doi:DOI:10.1109/ICPHM.2011.6024356

Delta Airlines. (2024). *Safety Management System*. Retrieved from Delta.com: https://www.delta.com/content/www/esg/en/2023/safety-management-system.html

Dou, X., & Tan, A. W. (2020, May). Big data and smart aviation information management system. *Cogent Business & Management*. doi:https://doi.org/10.1080/23311975.2020.1766736

EASA. (2023). *Management System Assessment Tool, Edition 2.* European Union (EU).

EASA. (2024a). *VIRTUA (Digital Transformation - Case Studies for Aviation Safety Standards - Virtualisation).* Retrieved from https://www.easa.europa.eu/en/downloads/140427/en

EASA. (2024b). *VIRTUA - Final dissemination event (webinar presentation).* Retrieved from https://www.easa.europa.eu/en/downloads/140427/en

Ellis, K. L., Prinzel, L. J., Davies, M. D., Krois, P. A., Mah, R. W., Oza, N. C., . . . Ackerson, J. R. (2022). *The In-time Aviation Safety Management System Concept for Part 135 Operations.* Hampton: NASA. Retrieved from https://ntrs.nasa.gov/api/citations/20220009144/downloads/DASC%20IASMS%20for%20Part%20135%20Final.pdf

Escobar, J. (2024). *Boeing and Bowtie in Safety Risk Management.* Skylibrary. Retrieved from https://skybrary.aero/sites/default/files/bookshelf/34780.pdf

EUROCONTROL. (n.d.). *Aeronautical Information Exchange Model*. Retrieved August 16, 2024, from https://aixm.aero/homepage

Evans, J. W., Cornford, S. L., Kotsifakis, D., Crumbley, T., & Feather, M. S. (n.d.). *Enabling Assurance in the MBSE.* NASA. Retrieved from https://core.ac.uk/reader/323104110

FAA. (2023, September 29). Order 8040.4C. *Safety Risk Management Policy*. Washington, DC.

FAA. (2024, September 13). AHCS & Hazard Analysis. Washington, DC.

Fan, S., & Yang, Z. (2023, May). Towards objective human performance measuremetn in maritime safety: A new psychophysiological data-driven machine learning model. *Reliability Engineering & System Safety, 233*. Retrieved from https://www.sciencedirect.com/science/article/pii/S0951832023000182

Farinha, D., Pereira, R., & Almedia, R. (2023). A framework to support robotic process automation. *Journal of Information Technology, 39*(1). doi:https://doi.org/10.1177/02683962231165066

Future Sky Safe. (2019). *Aviation Safety Dashboards.* European Commission. Retrieved from https://www.futuresky-safety.eu/wp-content/uploads/2019/07/FSS_White-paper_Aviation-Safety-Dashboards.pdf

Gilbert, T., Lamont, J., & Bell, D. A. (2024). *Survey of Performance-Based and Risk-Based Oversight Strategies and Methods.* Washington, DC.: Federal Aviation Administration (FAA).

Hammond, D. M., King, A. L., Joe, M., & Miller, J. R. (2023). Understanding the relationship between safety culture and safety performance indicators in U.S. nuclear waste cleanup operations. *Safety Science*, *166*. doi:https://doi.org/10.1016/j.ssci.2023.106241

Han, E. (2023, October 31). *Setting Business Goals & Objectives: 4 Considerations*. Retrieved from Harvard Business School: https://online.hbs.edu/blog/post/business-goals-and-objectives

Hart, L. E. (2015). *Introduction to Modle-Based System Engineering (MSBE) and SysML.* INCOSE. Retrieved from https://www.incose.org/docs/default-source/delaware-valley/mbse-overview-incose-30-july-2015.pdf

Healey, R. (2024). *The Importance of Data Lifecycle Management: What It Is And Why You Need It*. Retrieved from Formit: https://formiti.com/the-importance-of-data-lifecycle-management-what-it-is-and-why-you-need-it/

Hollinger, K. B. (2021, May). Evaluating Safety Management Effectiveness. MITRE.

Honeywell . (2020, August). *Honeywell Uses Blockchain To Digitize Aircraft Records, Parts Pedigree Data.* Retrieved from https://www.honeywell.com/us/en/press/2020/08/honeywell-uses-blockchain-to-digitize-aircraft-records-parts-pedigree-data

Honeywell. (2024). *Four Ways Blockchain will Change Aircraft Maintenance.* Retrieved from https://aerospace.honeywell.com/us/en/about-us/blogs/four-ways-blockchain-will-change-aircraft-maintenance

IATA. (2020, October 7). Airline Value Chain Reference Model. Retrieved from https://airtechzone.iata.org/aidm_model/24.1/index.htm?goto=2:1:1:68

IATA. (n.d.). *Aviation Info. Data Exchange (AIDX)*. Retrieved August 16, 2024, from https://www.iata.org/en/publications/info-data-exchange/

IATA. (n.d. 2). *Incident Data eXchange (IDX)*. Retrieved August 16, 2024, from https://www.iata.org/en/services/data/safety/gadm/idx/

IATA. (n.d. 3). *Airline Industry Data Model*. Retrieved August 16, 2024, from https://www.iata.org/en/about/corporate-structure/passenger-standards-conference/architecture-technology-strategy/industry-data-model/

IATA. (n.d.). Root Cause Analysis (virtual classroom course). Retrieved August 12, 2024, from https://www.iata.org/en/training/courses/root-cause-virtual/tals42/en/

ICAO. (n.d.). Retrieved September 2024, from USOAP Continuous Monitoring Approach (CMA): https://www.icao.int/safety/cmaforum/Pages/default.aspx

ICAO. (2021). *Big Data Analytics and Tools to Support Aviation Analysis.* Montreal: ICAO Webinar on Air Trasport Data and Analysis.

ICAO. (2022). Overview on Exchange Models (Working Paper). *Assembly - 41st Session, Technical Commission, Agenda Item 31: Aviation Safety and Air Navigation Standardization.* Retrieved from https://www.icao.int/Meetings/a41/Documents/WP/wp_507_en.pdf

ICAO. (2024, March). SM ICG Industry Engagement Group: Product Use Survey Results.

ICAO 9734. (2017). Safety Oversight Manual, 3rd Edition. *Part A - The Establishment and Management of a State Safety Oversight System*. Montreal, Quebec, Canada: ICAO.

ICAO 9753. (1999, September 14-24). Report of the Accident Investigation and Prevention (AIG) Divisional Meeting (1999). *Document 9753*. Montreal, Canada: ICAO.

ICAO 9859. (2018). *Safety Management Manual.* Montreal: International Aviation Civil Organization (ICAO).

ICAO Annex 19. (2016, July). Safety Management, 2nd Edition. Montreal, Quebec, Canada: ICAO.

ICAO Annex 6. (2018, July). Operation of Aircraft, 11th Edition. *Part I - International Commercial Air Transport - Aeroplanes*. Montreal, Quebec, Canada: ICAO.

ICAO SM ICG. (2019a). *Guidance for Comprehensive Safety Performance Management in a State Safety Programme.* SKYLibrary. Retrieved from http://bit.ly/SMICG

ICAO SM ICG. (2022, March). Safety Management Terminology.

IEEE. (2024). *IEEE Big Data*. Retrieved from https://bigdata.ieee.org/about

International Council on Systems Engineering. (2024, May 6). Systems Engineering Body of Knowledge (SEBoK) version 2.10. Retrieved from https://sebokwiki.org/wiki/Guide_to_the_Systems_Engineering_Body_of_Knowledge_(SEBoK)

ISO 19011. (2018, July). Guidelines for Auditing Management Systems, 3rd Edition. *ISO 19011:2018*. Geneva, Switzerland: International Standards Organization.

Iswarappa, A. J. (2015). A Brief Introduction on Big Data 5Vs Characteristics and Hadoop Technology. *International Conference on Intelligent Computing, Communication, & Convergence.* Odisha: Elsevier B.V. doi:10.1016/j.procs.2015.04.188

JALUX. (2022, April). *JALUX Inc. Announces Strategic Investment in Block Aero.* Retrieved from https://block.aero/2022/04/11/jalux-invests-in-block-aero-an-aviation-blockchain-company/

Janakiraman, V. M., & Nielsen, D. (2016). Anomaly detection in aviation data using extreme learning machines. *International Joint Conference on Neural Networks (IJCNN)*, 1993-2000.

Key, K., Hu, P. T., Choi, I., & Schroeder, D. J. (2023). *Safety Culture Assessment and Continuous Improvement in Aviation: A Literature Review.* Washington, DC: Federal Aviation Administration (FAA). Retrieved from https://www.faa.gov/about/initiatives/maintenance_hf/Safety_Culture_Assessment_Continuous_Improvement_in_Aviation.pdf

Key, K., Hu, P., Schroeder, D., & Choi, I. (2023). *User Handbook: FAA Maintenance Safety Culture Assessment and Imrovement Toolkit (M-SCAIT).* Washington, DC.: Federal Aviation Administration (FAA).

Kirwan, B. (2024). The Impact of Artificial Intellegence on Future Aviation Safety Culture. *Future Transportation*, 349-379. doi:https://doi.org/10.3390/futuretransp4020018

Kobi, J. (2024, May). Developing Dashboard Analytics and Visualization Tools for Effective Performance Management and Continuous Process Improvement. *International Journal of Innovative Science and Research Technology, 9*(5). doi:https://doi.org/10.38124/ijisrt/IJISRT24MAY1147

Kyriakos, K. (2020, October 19). *The Bowtie Methodology as a Meta-Analysis Tool for Aviation Accident Investigation Reports*. Retrieved from AviationPros: https://www.aviationpros.com/aircraft/article/21158951/the-bowtie-methodology-as-a-meta-analysis-tool-for-aviation-accident-investigation-reports

L. F. F. M. Santos, e. a. (2022). Blockchain Information Based Systems in Aviation: The Advantages for Aircraft Records Management. *International Conference on Control, Automation and Diagnosis (ICCAD).* Lisbon, Portugal: IEEE.

Lamine, E., Thabet, R., Sienou, A., Bork, D., Fontanili, F., & Pingaud, H. (2020, May). BPRIM: An integrated framework for business process management and risk management. *Computers in Industry, 117*. Retrieved from https://www.sciencedirect.com/science/article/abs/pii/S0166361520300890

MA Software Solutions. (2024, March 23). SMS Suite Case Study: Spirit Airlines. St. Petersburg, FL.

National Academy of Engineering. (2004). *Accident Precursor Analysis and Management: Reducing Technological Risk Through Diligence.* National Academy of Engineering.

National Business Aviation Association (NBAA). (2021, July/August). Machine Learning Enhances SMS. *Business Aviation Insider*. Retrieved from https://nbaa.org/news/business-aviation-insider/2021-07/machine-learning-enhances-sms/

Odarchenko, R., Hassan, Z., & Zaman, A. (2019). *Use of Big Data in Aviation.* Research Gate. doi:10.4018/978-1-5225-7709-6.ch017

Oneto, L., Buselli, I., Dambra, C., Gallego, C., Martinez, M. G., Smoker, A., . . . Pejovic, T. (2021). Natural Language Processing and Data-Driven Methods for Aviation Safety and Resilience: from Extant Knowledge to Potential Precursors. *SESAR Innovation Days.* SESAR. Retrieved from https://www.sesarju.eu/sites/default/files/documents/sid/2021/papers/SIDs_2021_paper_50.pdf

Quatrini, E., Contantino, F., & Patriarca, R. (2020, July). Machine learning for anomaly detection and process phase classification to improve safety and maintenance activities. *Journal of manufacturing systems*. doi:10.1016/j.jmsy.2020.05.013

Rail Safety and Standards Board. (2024). *Bowtie Hub*. Retrieved from Rail Safety and Standards Board: https://www.rssb.co.uk/en/safety-and-health/guidance-and-good-practice/bowties

Reijers, H. (2021, January 30). Business Process Management: The evolution of a discipline. *Elsevier*. Retrieved August 2024, from Elsevier: https://pdf.sciencedirectassets.com/271439/1-s2.0-S0166361521X00024/1-s2.0-S0166361521000117/main.pdf?X-Amz-Security-Token=IQoJb3JpZ2luX2VjEBkaCXVzLWVhc3QtMSJHMEUCIQD1kRf%2BbYDiZAERpXFTRfJvt9U8K7567odlxnK0%2F%2B2u6wIgFNFNd5pE9r9f3DxK5LB4jzwYP7bijvxK9PBrUy

Reniers, G., Ponnet, K., & Van Nunen, K. (2022). Measuring Safety Culture Using an Integrative Approach: The Development of a Comprehensive Conceptual Framework and an Applied Safety Culture Assessment Instrument. *International Journal of Environmental Research and Public Health*.

Rizvi, A., & Srivastava, D. (2023). Exploring the Potentials of Robotic Process Automation: A Review. *Journal of Infomatics Electrical and Electronics Engineering (JIEEE), 04*(02, No. 100), 1-12. doi:https://doi.org/10.54060/jieee.2023.100

Skybrary. (n.d.). *ICAO Universal Safety Oversight Audit Programme (USOAP)*. Retrieved from Skybrary: https://skybrary.aero/articles/icao-universal-safety-oversight-audit-programme-usoap

Snowflake. (2024). *Top Five Data Science and Generative AI Trends for 2024.* Snowflake.

Solanki, U., Mehta, K., & Shukla, V. (2024). Robotic Process Automation and Audit Quality: A Comprehensive Analysis. *2024 11th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO),* (pp. 1-4). Noida. doi:10.1109/ICRITO61523.2024.10522375

Stephans, R. (2022). *System Safety for the 21st Century.* Wiley.

Systems Planning & Analysis (SPA). (2019). *Developing Effective MBSE Models in Digital Engineering Environments*. Retrieved from SPA: https://spa.com/news-insights/developing-effective-mbse-models-in-digital-engineering-environments/

Tambare, P., Meshram, C., Lee, C.-C., Ramteke, R. J., & Imoize, A. L. (2018). Performance Measurement System and Quality Management in Data-Driven Industry 4.0: A Review. *Sensor, 22(1)*. doi:https://doi.org/10.3390/s22010224

Tvaryanas, A. (2024). *Concept for a State Safety Program Medical Fitness Risk Management Reference Model.* FAA.

UK CAA. (2019, April 30). *CAP1760: Effective Problem Solving and Root Cause Identification (version 2)*. Retrieved from https://www.caa.co.uk/our-work/publications/documents/content/cap1760/

UK CAA. (2024). *https://www.caa.co.uk/safety-initiatives-and-resources/working-with-industry/bowtie/bowtie-templates/how-to-develop-a-bowtie-risk-assessment-model/*. Retrieved from CAA: https://www.caa.co.uk/safety-initiatives-and-resources/working-with-industry/bowtie/bowtie-templates/how-to-develop-a-bowtie-risk-assessment-model/

UK CAA. (n.d.). *CAA Gap Analysis Tool.* United Kingdom (UK) Civil Aviation Authority (CAA).

UK CAA Safety and Airspace Reguation Group. (2018). *Management of Change Evaluation Guide (SMS-OV-GU-001).* United Kingdom Civil Aviation Authority (CAA).

UK Department for Transport. (2023a, July 17). *Call for evidence outcome: The effectiveness and efficiency of the Civil Aviation Authority (CAA)*. Retrieved from https://www.gov.uk/government/calls-for-evidence/the-effectiveness-and-efficiency-of-the-civil-aviation-authority-caa/the-effectiveness-and-efficiency-of-the-civil-aviation-authority-caa

UK Department for Transport. (2023b, July 17). *Civil Aviation Authority review: report*. Retrieved from https://www.gov.uk/government/publications/civil-aviation-authority-public-body-review-terms-of-reference/civil-aviation-authority-review-report

United Airlines. (2024). *Safety in Action*. Retrieved from Delta.com: https://crreport.united.com/safety/safety-in-action

US Department of Energy. (2009). *Human Performance Improvement Handbook Vol 2: Human Performance Tools for Individuals, Work Teams, and Management.* Washington, DC: US DOE.

Vestola, E. N., & Eriksson, P. E. (2023). Engineered and emerged collaboration: vicious and virtuous cycles. *Construction Management and Economics, 41, NO 1*, 79-96. doi:https://doi.org/10.1080/01446193.2022.2140815

Visure. (2024). *Model-Based Systems Engineering Complete Guide*. Retrieved from Visure: https://visuresolutions.com/mbse-guide/best-mbse-tools-and-solutions/

Wang, Q., & Liu, Y. (2023, December). Blockchain for public safety: A survey of techniques and applications. *Journal of Safety Science and Resilience*, *4*(4), 389-395. doi:https://doi.org/10.1016/j.jnlssr.2023.09.001

Yaga, D., Mell, P., Roby, N., & Scarfone, K. (2018). *IR 8202 Blockchain Technology Overview*. National Institute of Standards and Technology (NIST). Washington, DC: US Department of Commerce. doi:https://doi.org/10.6028/NIST.IR.8208