ACHIEVING GLOBAL SAFETY DATA AGGREGATION AND PROCESSING:

The role of standards and data sharing platforms

Eugene Mwendwa, The MITRE Corporation

Introduction

> **"Aviation safety is a team effort."**
>
> *—Michael Huerta, former FAA Administrator (2013–2018)*

The need for global cooperation in aviation safety was evident on December 7, 1944, at the Chicago Convention, when 52 nations gathered and established the Provisional International Civil Aviation Organization.[1] Two years later, the agreement was ratified, and the International Civil Aviation Organization (ICAO) was formed. In that agreement, the signees agreed to pursue "certain principles and arrangements in order that international civil aviation may be developed in a ***safe*** and orderly manner and that international air transport services may be established on the basis of equality of opportunity and ***operated soundly*** and economically."[2] The fundamental strategic objective of ICAO to improve safety in civil aviation worldwide has not changed in the intervening 75 years.

Through technological advancements in fourth-generation aircraft[3] and air traffic management, along with regulations, programs like the Commercial Aviation Safety Team (CAST) contributed to an 83 percent reduction in the domestic commercial aviation fatality risk between 1998 and 2008.[4] As Michael Huerta said when he served as Administrator of the Federal Aviation Administration (FAA), "Data and collaboration are the backbones of aviation safety, and ASIAS [Aviation Safety Information Analysis and Sharing] has directly contributed to our impressive commercial aviation safety record in the United States." More recently, FAA Administrator Stephen Dickson noted the potential of ASIAS for further contributions to aviation safety. "ASIAS is one of the crown jewels of the aviation system in the United States. It is … an example of the kind of collaboration and safety innovation we can use to lead the global aviation safety system to even higher levels of performance," he said. The efforts of CAST and ASIAS have and continue to contribute to a near-zero fatality rate in commercial aviation in the United States. Through data collection, sharing, collaboration, and a reporting culture that is built on trust and data protections, all stakeholders (government, manufacturers, operators, and employee groups) have shifted from assessments of accidents and their precursors to identification, assessment, and mitigation of emerging risks. More recently, in the 2020–2022 edition of ICAO's Global Aviation Safety Plan, ICAO set a goal of achieving and maintaining zero fatalities in commercial aviation by 2030

1 https://www.icao.int/publications/pages/doc7300.aspx

2 https://www.icao.int/publications/pages/doc7300.aspx

3 https://www.skybrary.aero/bookshelf/books/4009.pdf. Aircraft with advanced fly-by-wire technology and flight envelope protection capabilities.

4 https://www.skybrary.aero/bookshelf/books/3511.pdf

and beyond, and continuously enhancing safety performance internationally.

Meanwhile, new entrants—from unmanned aircraft systems (UAS) to commercial space vehicles to urban air mobility vehicles—are coming on the scene. These new vehicles and their novel missions create a much more complex aviation environment calling for new approaches to safety management.

Are we ready on a global scale to monitor flight safety, to integrate UAS and other emerging entrants, and to provide effective and actionable safety intelligence? Will current data and existing curation processes, aggregation, and analytical methods enable us to prevent future accidents? Currently, the answer to each of those questions is no, so there is much work to be done.

While the reduction in flights during the COVID-19 pandemic helped to reduce total accidents and fatalities in 2020, air traffic is already starting to return. According to the latest forecasts by the International Air Transport Association, worldwide passenger traffic is projected to reach pre-pandemic levels sometime in 2023—just two years from now—and will continue to grow at a rate of nearly 4 percent annually.[5] The pandemic not only affected traffic levels but also introduced new kinds of risk due to shifting demands, such as deficiency in pilot proficiency, impacts to airframes from prolonged storage, and shifting business models. With the return of traffic, so too will return the more complex safety environment that existed before the pandemic.

So, what is needed for the world to achieve the safety goals ICAO has laid out? For starters, we need to continually improve our understanding of the risk factors that may lead to accidents. Much has been achieved in this area, but in order to take this understanding to the next level, researchers around the world need to have a common view of global safety data and information and collaborate to make the best use of it.

Achieving zero fatalities by 2030 will require global cooperation and participation in two technical areas:

1. Open data standards
2. Data-sharing analytic platforms

Open Data Standards

> **“**
>
> **Open Standards” are standards made available to the general public and are developed (or approved) and maintained via a collaborative and consensus driven process. “Open Standards” facilitate interoperability and data exchange among different products or services and are intended for widespread adoption.**
>
> **”**
>
> *—International Telecommunication Union (United Nations specialized agency)*[6]

5 https://www.iata.org/en/iata-repository/publications/economic-reports/an-almost-full-recovery-of-air-travel-in-prospect/

6 “Definitions of Open Standards,” *ITU 2021.* Retrieved 17 August 2021. https://www.itu.int/en/ITU-T/ipr/Pages/open.aspx

With the increase in data available across high-reliability organizations and an increased focus on data-driven decision making to achieve organizational objectives, the need to share standardized and validated data has never been greater. In addition, international, national, and inter-organizational data sharing requires common data standards to facilitate connections across agencies as they monitor key performance indicators and conduct collaborative research and safety mitigation efforts. Currently, addressing data quality issues consumes 80 percent of data scientists' time. Open data standards can help address this challenge.[7]

A set of common and extensible data standards is the backbone of any next-generation aviation safety system and will likely need to evolve via open collaboration between government, industry, academia, and other stakeholders. Well-defined data standards will facilitate communication across heterogeneous systems and platforms, reducing data processing overhead and shifting the focus to collaboration and new modes of centralized and distributed analysis. In all, this will strengthen the platform and the resulting analysis by reducing variability. The net effect will be an increase in information sharing (volume, velocity) and improvement in the quality of results (veracity), with a larger availability of datapoints (variety) for data-driven decisions (value), ultimately enabling safety intelligence.[8] [9]

The call for adherence to data standards is being voiced not only by the FAA but across a range of industries (see Federal Data Strategy, 2020[10]) where data-driven decision making and accountability are seen as critical to the advancement of safety initiatives. MITRE, as a data steward for many proprietary and non-proprietary data sources, is in a unique position to develop and promote validated best practices that support the adherence to data standards.

Achieving true data aggregation and processing globally will require adopting data standards to promote interoperability across States and safety systems. Several well-known standards have been created by traditional standard development working groups to promote information-sharing initiatives for commercial aviation—including the CAST/ICAO Common Taxonomies, the Human Factors Analysis and Classification System, and the ASIAS Aviation Safety Action Program's taxonomy. These well-known standards are developed and adopted through a consensus-based decision-making process with subject matter experts following a formal process. While thorough, that process can be time-consuming. To keep pace with today's rapidly evolving environment, with its increasingly diverse set of National Airspace System users, a more agile approach is needed. In contrast to traditional standards development processes, open data standards developed by agile principles incorporate a less formal process focused on being responsive to the rapidly changing aviation ecosystem.

Open data standards are needed to manage the constantly growing and evolving aviation data. Open data standards are vital to the interoperability, reusability, discoverability,

7 Pres, G. (2016, March 3). *Cleaning Big Data*. Retrieved from http://www.forbes.com

8 Advances in Computational Intelligence and Communication Technology, Xiao-Zhi Gao, Shailesh Tiwari, Munesh C. Trivedi, Krishn K. Mishra, page 410, Springer Singapore, 2021, doi 10.1007/978-981-15-1275-9

9 "The 5 V's of big data". *Watson Health Perspectives*. 17 September 2016. Retrieved 20 January 2021.

10 Federal Data Strategy. Action 19: Develop Data Quality Measuring and Reporting Guidance

comparability, and consistency of analyses. Most importantly, these standards make safety data sets fluid and promote common understanding, sharing, and use of better quality data. With open data standards, complex connections can be made across diverse safety data sets powered by secure, scalable data sharing analytic platforms like Collaborative Research Environments (CREs) and Joint Learning Platforms (JLPs). These are discussed in the next section.

Data Sharing Analytic Platforms

Collaborative Research Environments

CREs are a mechanism to further expand data collection from regulated entities across the globe, while keeping data secure. CREs also enable the sharing of data analyses, metrics, and insights, and are an extension of traditional data infrastructure and analytic platforms, which are enclosed by brick and mortar and private networks. A CRE is a managed, secure, and integrated virtual shared environment where entities can bring data and jointly develop complex connections and discover safety insights. This virtual extension, enabled by explicit application programming interfaces (APIs) for the exchange of data and models, is essential to accommodate collaboration and generic tooling to support input from multiple safety teams. Despite being technically diverse and geographically dispersed, a CRE can enable approved entities to jointly develop open standards, to understand dependencies and connections, reuse and enhance advanced predictive models, and streamline delivery of research activities into operations. The secure environment offers a common data management platform with simplified access to best-practice analytical tools, common data, vetted data-blending algorithms, and modeling resources so that safety practitioners can focus more on solving complex problems and less on configuring the software, creating analytical workflows, and debugging data quality issues. This is a space where open data standards are vital for success.

As illustrated in Figure 1-1, the power of a CRE lies in its ability to bring together information and context not available to any single stakeholder. Stakeholder data is generally limited to a single perspective; for example, pilot reports and onboard flight data instrumentation. However, this data lacks the context of neighboring traffic, routing and sequencing congestion, convective weather patterns, controller-pilot voice data, and other factors that influence safe and efficient flight operations. Adding this context requires the acquisition and fusion of additional surveillance, weather, and traffic management systems data. The workload to process this information globally requires a secure data infrastructure with scalable compute resources.

However, since many States view safety data sets as sensitive, enabling true worldwide data aggregation using a CRE may not be realistic. Especially where the airspace is operated by the State, or where data sharing, just culture, and data protections are not as mature, one technological solution to enable global collaboration would be to architect a system in a way that enables States to learn jointly without the need to transfer their data. JLPs, as described in the next section, are one potential solution to this challenge.

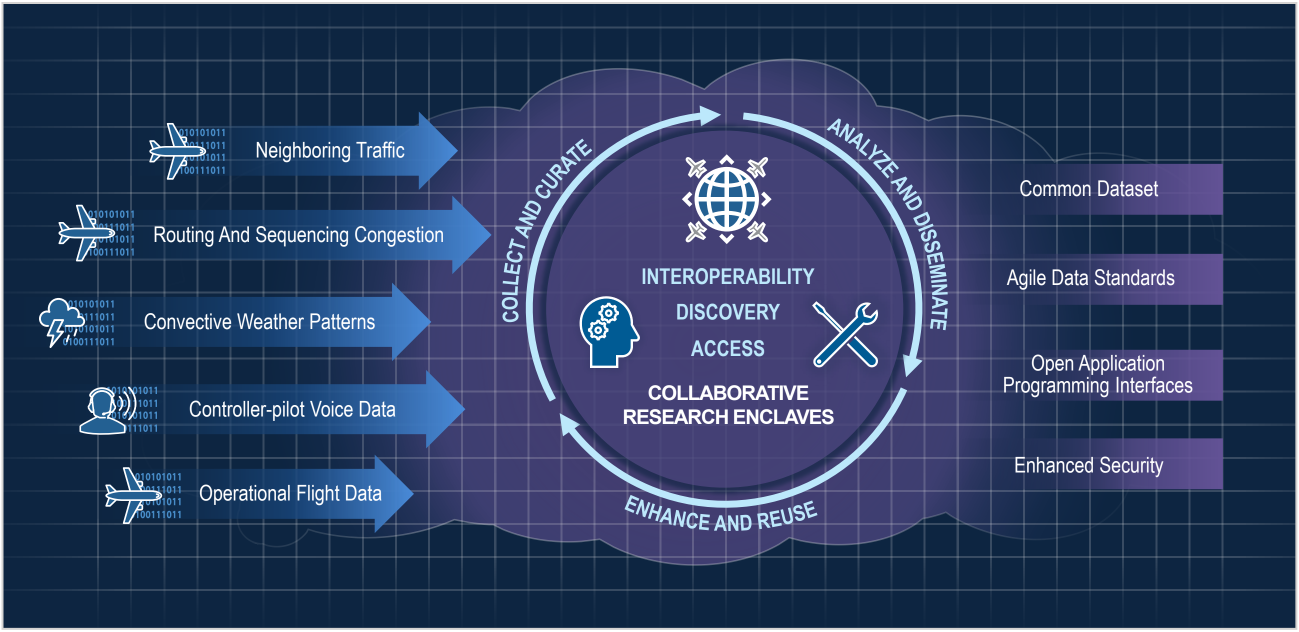

FIGURE 1-1: COLLABORATIVE RESEARCH ENVIRONMENTS

Joint Learning Platforms

JLPs utilize an emerging machine learning technique (federated learning) that meets the challenges of decentralized, sensitive data[12] and, as in the case of CREs, will benefit greatly from the use of open data standards. "Federated learning is a machine learning setting where multiple entities (clients) collaborate in solving a machine learning problem, under the coordination of a central server or service provider. Each client's raw data is stored locally and not exchanged or transferred; instead, focused updates intended for immediate aggregation are used to achieve the learning objective."[13] As illustrated in Figure 1-2, an entity's original raw data is never moved from its original location. Each joint learning node (Nation 1, Nation 3, Operator, and Manufacturer) analyzes its data locally. A global model is created, and duplicate versions of that model are sent to each distributed data node. The duplicate local model learns from its data set and sends back the local analysis results. The analysis results are synthesized with the analyses from other entities and integrated into the global model, which is operated by a trusted aggregation broker. This process repeats at a frequency, coordinated by the global model, and constantly refines and improves the global model. Joint learning enables machine learning use cases that were previously impossible due to data sensitivity.

Despite its recent emergence, joint learning has already been successfully deployed by industries across the globe, and it could have been a valuable tool to identify issues in advance of the recent Boeing 737 MAX accidents. Data sovereignty and

[12] H. B. McMahan, E. Moore, D. Ramage, and B. A. y Arcas, "Federated Learning of Deep Networks using Model Averaging," *CoRR*, vol. abs/1602.05629, 2016

[13] P. Kairouz *et al*., "Advances and Open Problems in Federated Learning," *CoRR*, vol. abs/1912.04977, 2019.

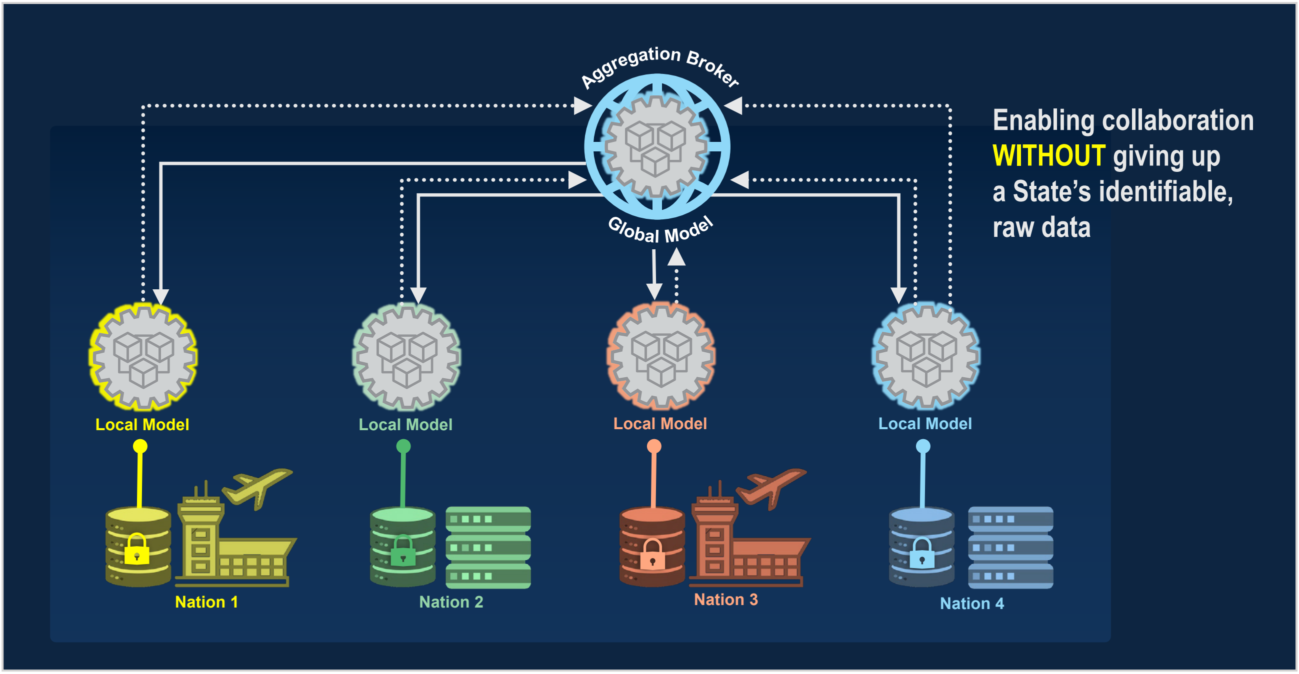

FIGURE 1-2: JOINT LEARNING PLATFORMS

differential privacy enabled by federated learning would likely have allowed a partnership to build a shared model to discover anomalies. The Boeing 737 MAX was newly certified for airworthiness, but each country, airline, and manufacturer had access to only a small number of flights from which to identify anomalies. States and airlines are reluctant to share flight data with each other or with a trusted third party for use in aviation safety models. Incorporating more flights, flight hours, and miles would have produced a larger data set with which to detect potential anomalies and predict safety issues.

Although potentially powerful, as with any new technology, there are challenges that must be overcome to enable the full potential of joint learning to enhance safety analysis globally.

Problems in joint learning are highly interdisciplinary and exist at the intersections of multiple fields, such as machine learning, distributed optimization, cryptography, security, fairness, compressed sensing, distributed systems, information theory, and statistics. In particular, the lack of centralized data exacerbates existing challenges with data. Data may be unbalanced, with larger entities producing an overwhelming plurality of data that may obscure potential insights and limit usefulness for smaller entities. This is challenging to deal with in typical analysis and machine learning problems, but lack of access to raw data can make even detection of such imbalances a challenge for joint learning. These challenges also extend to verifying data quality, repeated analysis on evolving data, and debugging and interpretability of models. Care must be taken to ensure that bias in training data is accounted for in models and that model results are fair, without exposing sensitive attributes. Joint learning approaches must also carefully consider the tension between privacy and robustness. Enhancing privacy guarantees can reduce the robustness and ultimate usefulness of the model.[14] This is an ongoing area of research; successful outcomes will require academia and artificial intelligence industry working groups to contribute to open-source joint learning services that advance optimization models in heterogeneous environments and benchmarking and evaluation frameworks.

Non-Technical Success Factors for Data Sharing and Analytics

While technical solutions such as CREs and JLPs exist to standardize and protect data, a number of non-technical hurdles must be overcome for successful data sharing and analytics. These include building and maintaining trust and developing safety intelligence capabilities.

Positive just culture built on trust is the glue to adoption of CREs and JLPs internationally. The trust that the U.S. aviation community has established was decades in the making. Data-sharing programs will need to make trust paramount, embrace a non-punitive safety culture, and be bound by the governing principles of safety management and information protection. Ensuring the proper legal protections of the data and information required to enable data-driven decision making is fundamental and must be adhered to by all stakeholders—at regional, state, government, and corporate levels.[15]

As we move into more timely sharing of safety intelligence from operational data—like Operational Flight Data Monitoring, also known as Flight Operations Quality Assurance—that could inform on a path to a potential accident, laws governing the use of safety data and safety intelligence will have to be reexamined at the international and national levels. National laws and regulations need to take into consideration the level of sensitivity and identity of data, as well as open data standards, while exploring more liberal protocols to effectively share safety intelligence globally.

Safety intelligence is curated information that feeds the processes supporting data-driven decision making and the timely mitigation of safety risks. For an organization to acquire safety intelligence, an investment must be made in establishing processes that generate actionable, reliable, and timely information that is then integrated into the organization's business processes.

The safety intelligence of an organization is characterized by its ability to generate and transform data into actionable information to support evidence- and risk-based decision making. Safety data must be curated and integrated into an organization's existing systems to have meaningful operational impact. Safety intelligence requires safety data to meet the following criteria:

- Actionable – fulfills defined need for insight into safety risk
- Integrated – supports capabilities that are embedded in an established operational process
- Reliable – trusted information supported by measured accuracy
- Timely – delivered in time to enable mitigation of identified risk
- Curated – vetted, customized source of verifiable information

An organization that can deliver the right data at the right time to a knowledgeable workforce that is empowered to use the information to enhance their decision making is supporting the principle of safety intelligence generation.

[14] P. Kairouz *et al.*, "Advances and Open Problems in Federated Learning," *CoRR*, vol. abs/1912.04977, 2019.

[15] https://flightsafety.org/wp-content/uploads/2017/02/FSF-GSIP-Information-Protection-Toolkit_v4.pdf

What's Needed?

To drive to zero passenger fatalities and maintain a fatality-free record worldwide, technology solutions like CREs and JLPs can be leveraged to enable timely data collection and processing. This will require the following:

- A CRE pilot should be accelerated, with an international entity utilizing multiple global aviation data sources and analytics to provide insights into aviation safety challenges. With a managed and curated data set in a CRE, international standards working groups could apply agile principles to develop open safety data standards and APIs. They would also define a new standard of operational safety data protection that promotes cross-border data sharing and aggregation.
- The FAA should play a leading role in establishing a JLP with trusted international partners. This will pave the way to understanding how to manage JLPs at scale and will promote early adoption.
- The FAA, working with international aviation bodies, should lead the development of a model or framework that informs the global community on the appropriate protections for effective sharing—not so strict that there is no data sharing, but not so lax that there is the potential for misuse of sensitive or proprietary safety data.

While data and collaboration continue to be the backbones of aviation safety, open data standards and data analytic platforms such as CREs and JLPs can take aviation safety to the next level. They will provide a common view of global safety data to all stakeholders, facilitate connections across organizations for the collective monitoring of key performance indicators, and support collaborative research and the implementation of safety mitigation efforts. Trust-building and the development of safety intelligence capabilities will further enhance these opportunities. By pursuing these goals together, the global aviation community will be well-positioned to achieve the ICAO safety goals for the next decade. With collaboration and the new tools available to the global aviation community, zero fatalities in commercial aviation by 2030 and the continuous enhancement of safety performance internationally are real possibilities.

Bibliography

H. B. McMahan, E. Moore, D. Ramage, and B. A. y Arcas, “Federated Learning of Deep Networks using Model Averaging,” *CoRR*, vol. abs/1602.05629, 2016.

P. Kairouz et al., “Advances and Open Problems in Federated Learning,” *CoRR*, vol. abs/1912.04977, 2019.

Acronyms

API	Application Programming Interface
ASIAS	Aviation Safety Information Analysis and Sharing
CAST	Commercial Aviation Safety Team
CRE	Collaborative Research Environment
FAA	Federal Aviation Administration
ICAO	International Civil Aviation Organization
IDC	International Data Corporation
JLP	Joint Learning Platform
UAS	Unmanned Aircraft Systems

About MITRE

MITRE's mission-driven teams are dedicated to solving problems for a safer world. Through our public-private partnerships and federally funded R&D centers, we work across government and in partnership with industry to tackle challenges to the safety, stability, and well-being of our nation.

The views, opinions, and/or findings contained herein are those of the author(s) and should not be construed as an official government position, policy, or decision unless designated by other documentation.

MITRE | SOLVING PROBLEMS FOR A SAFER WORLD®