IBM Center for
The Business of Government

Improving Performance Series

Incident Reporting Systems

Lessons from the Federal Aviation Administration's Air Traffic Organization

Russell W. Mills
Bowling Green State University

Improving Performance Series 2013

Incident Reporting Systems: Lessons from the Federal Aviation Administration's Air Traffic Organization

Russell W. Mills
Assistant Professor
Department of Political Science
Bowling Green State University

IBM Center for
The Business of Government

Table of Contents

Foreword

Daniel J. Chenok

William M. Takis

On behalf of the IBM Center for The Business of Government, we are pleased to present this report, *Incident Reporting Systems: Lessons from the Federal Aviation Administration's Air Traffic Organization,* by Russell W. Mills, Bowling Green State University.

Incident reporting systems are an integral part of many agencies' operations. For example, the Veterans Health Administration collects data on incidences of errors made during surgeries, the Food Safety and Inspection Service collects data on incidences of errors in meat inspection plants, and the Occupational Health and Safety Administration collects data on incidences of work-place injuries.

But collecting the raw numbers of when an incident occurs does not necessarily help prevent future incidents from happening. Agency managers need to analyze data at a much finer level to understand why incidents occur and what can be done to prevent them in the future. This is often called the black box of performance management—understanding the relationships that connect potential changes in operations to the needed improvements in outputs and outcomes being measured.

In this report, Dr. Mills narrates a case study of the Federal Aviation Administration's Air Traffic Organization (ATO). He unpacks the black box of how it developed, used, and refined its incident reporting systems over the past few years. The ATO is the home agency for over 16,000 air traffic controllers who manage 7,000 aircraft flying over the U.S. at any point in time. They are responsible for maintaining safe separation between flying aircraft in order to reduce the risk of collisions. If aircraft violate separation standards, these incidents are called "loss of standard separation" and are reported as "operational errors."

Dr. Mills describes how the ATO's incident reporting systems evolved since the late 1990s with the introduction of voluntary self-reporting of errors by air traffic controllers and the use of increasingly sophisticated electronic tracking equipment. He describes the multiple systems used to track operational errors—some qualitative and some quantitative—that interact to provide a rich picture of air traffic operations. The report also

describes the various analytic teams that comb over the incident reports at the local level several times a week to reveal patterns. This increased internal performance information allowed more near-real time analysis of patterns of errors so they could be detected and fixed.

However, he writes that, ironically, this better data collection initially alarmed external stakeholders—the traveling public and Congress. To them, it seemed that there was a dramatic increase in the number of operational errors. In fact, the increased reporting of incidents that had previously been undetected or unreported led to a greater understanding of trends and causal factors, thereby allowing ATO to put in place corrective actions. While this led to a safer air traffic system, it created political problems for the agency.

Dr. Mills reports that ATO overcame these political problems by creating a new risk-based reporting system for the traveling public and Congress that demonstrated that the new elements of its incident reporting systems are contributing to greater safety. Based on the experience of ATO's evolving incident report systems, Dr. Mills offers a set of strategic, management, and analytical lessons that could be applied by other agencies that may also be in the process of increasing the sophistication of their own incident reporting systems.

We hope you find this report useful in understanding the technical and political dynamics associated with the implementation of, or changes in, an incident reporting system and how to better manage the understanding and expectations of internal and external stakeholders in the process.

Daniel J. Chenok
Executive Director
IBM Center for The Business of Government
chenokd@us.ibm.com

William M. Takis
Vice President
U.S. Federal Civilian Industry Leader
bill.takis@us.ibm.com

Executive Summary

New data and information sources that measure agency performance are a major topic of discussion within the performance management community. Executives and managers in a variety of agencies have spent valuable resources collecting and analyzing data on how effectively employees are implementing the organization's mission. In many agencies, the outputs and outcome of the work of frontline employees are difficult or impossible to observe without innovative data collection techniques.

The focus on collecting data is particularly important in high-reliability agencies, where poor performance can result in high-impact consequences such as a loss of life or large-scale disasters. Among the new sources of data used by many high-reliability agencies to measure performance are voluntary incident disclosures by employees and contractors, along with electronic incident or error collection conducted by frontline employees in operational or procedural agencies.

Voluntary incident reporting programs encourage employees to report instances of non-compliance or violations of agency procedures in exchange for confidentiality and protection from punitive action. These reporting programs also give the agency frontline feedback on the hazards and risks of current operations, and provide an accurate account of how well the agency is performing its critical tasks. However, because of the highly sensitive and subjective nature of voluntarily submitted data, using this valuable resource to measure agency performance is complicated.

High-reliability agencies have invested in new technologies to electronically collect data on the operational performance of employees. However, increased reporting of incidents can often be misinterpreted by agency leadership, unions, and members of Congress and can lead to increased oversight and public scrutiny (Mills 2010). This puts growing pressure on managers in high-reliability agencies to design performance indicators that can be used to honestly assess agency performance, while also addressing the concerns of both internal and external stakeholders.

The purpose of this report is to explore the challenges and opportunities associated with the use of performance indicators in high-reliability organizations. Specifically, what concerns are important for managers and executives to consider before adopting and implementing new performance indicators that measure agency outputs and outcomes through the use of incident and error reporting systems? Additionally, what strategies are available to executives and managers in agencies, particularly high-reliability agencies, to design performance indicators that use these sources of data while also addressing the concerns of both internal and external stakeholders?

These questions are addressed by presenting a rich case study that examines the evolution of safety performance indicators used by the Federal Aviation Administration's Air Traffic

Organization (ATO). The report identifies strategic, managerial, and analytical lessons learned from the FAA ATO's experience in using incident reporting systems to develop performance indicators:

Strategic Lessons

- As agencies report more performance information, including incident reporting systems, there will be increased scrutiny of agency performance from external stakeholders.
- Agencies must proactively educate key stakeholder groups on new reported performance indicators and incident reporting systems.
- Agencies must develop indicators that are easily understood by external stakeholders.

Management Lessons

- In order to be used by agency leaders for decision-making, data analytics processes must focus on reviewing performance information derived from incident reporting systems in a timely manner.
- Collaboration among managers and employees on the use of sensitive incident reporting performance information is critical to success.

Analytical Lessons

- Agencies must balance the pressure to develop indicators in a timely manner with the need for rigorous indicators that effectively measure agency performance.
- Agencies need to develop and invest in rigorous analytical techniques to make the most of new sources of performance information from incident reporting systems.

Introduction: Measuring Risk in High-Reliability Organizations

A great paradox of the performance measurement movement is that as agencies use innovative, rigorous methods to collect and report more information on mission achievement than ever before, they are often subject to much more scrutiny from external stakeholders, including the media and members of Congress, who lack a framework for interpreting or understanding the new sources of data. This is particularly true in agencies where poor performance can result in high-impact consequences, such as loss of life or large-scale disaster. In these high-reliability agencies, executives and managers must develop strategies to both rigorously and accurately measure performance in ways that can be used internally to avoid a large-scale failure, while also using that information to build external support for the agency's capacity to achieve its mission.

The Federal Aviation Administration (FAA) is a high-reliability agency that has often come under intense scrutiny from external stakeholders due to a lack of understanding of performance information. The FAA is the primary regulatory agency in charge of air transportation in the United States. In addition to providing safety oversight of air carriers, pilots, and aircraft operating in the United States, the FAA also serves as the operator of the nation's air traffic control system through its Air Traffic Organization (ATO).

A major responsibility of ATO air traffic controllers is to give pilots instructions ensuring that aircraft meet FAA standards of safe separation. However, loss of standard separation (LoSS) between aircraft routinely occurs in the national airspace system (NAS) due to pilot deviations, air traffic controller error, or outdated or inaccurate published procedures. LoSS due to air traffic controller errors represents a visible and robust indicator of ATO's performance in maintaining safety. As is common in many high-reliability organizations, a high-risk LoSS, such as a midair collision, would result in a catastrophic event. However, most reported incidents of LoSS are not serious and provide an opportunity for organizational learning and performance improvement to prevent more serious events.

Traditionally, the ATO had measured its performance in enforcing separation standards through the number of overall operational errors (OEs) committed by air traffic controllers. However, this safety performance indicator was problematic from a number of perspectives, including that the raw number of OEs in itself did not allow the ATO to effectively learn from the errors to develop corrective actions. Additionally, there was often a disincentive for reporting by controllers, who could be subject to punitive action, and by facility managers, whose raises and bonuses were tied to reducing the number of OEs from year to year.

To improve both reporting and the agency's ability to use the information contained in reports, the ATO implemented a voluntary reporting system known as the Air Traffic Safety Action Program (ATSAP) to encourage more detailed and frequent reporting of safety hazards by frontline controllers in exchange for immunity from punitive action. The ATO has also invested significant resources in developing and deploying an electronic system called the Traffic

High-Reliability Organizations

- A high-reliability organization (HRO) is an organization that must succeed in avoiding catastrophes in an environment where normal accidents can be expected due to risk factors and complexity.
- Characteristics of high-reliability organizations
  - — Operate in environment of high complexity and risk where normal accidents (Perrow 1984) occur that may lead to catastrophic events.
  - — Use incident reporting systems where employees report, or equipment detects, incidents or errors committed by employees which are then analyzed for cause and used to develop corrective actions.
  - — Using reports of small errors, HROs engage in organizational learning by assessing the risk of each error, determining the root cause, and implementing a corrective action to prevent or mitigate future incidents.
- Examples of high-reliability organizations in the federal government
  - — National Regulatory Commission
  - — United States Department of Agriculture's Food Safety and Inspection Service
  - — United States Navy's Nuclear Submarine Forces
  - — Department of Energy Laboratories
  - —Federal Aviation Administration's Air Traffic Organization

Analysis Reporting Program (TARP), which automatically detects LoSS between two aircraft. The implementation of ATSAP and TARP has given the ATO access to valuable information on the areas of highest risk in its operation, as well as data on the performance of its employees.

However, the ATO's increased access to performance information has also resulted in increased scrutiny of its performance indicators by members of Congress, the media, and key interest groups. As a recent news account noted:

> But over the past year or so, agency leaders have been mulling how to describe, categorize and learn from this barrage of additional information. The FAA, according to some industry officials, also has been concerned about the likely reaction on Capitol Hill, where some congressional leaders have been critical of earlier jumps in the frequency of midair close calls (Pasztor and Nicas 2013).

In September 2013, the FAA released its first analysis of ATO safety information containing both ATSAP and TARP reports. The overall number of validated LoSS jumped from 1,895 in FY 2011 to 4,394 in FY 2012 (FAA 2013), due in part to increased reporting of incidents through TARP and ATSAP. Some observers have reacted by concluding that the skies are less safe because the number of reported incidents of LoSS has "shot up" from 2012 to 2013 (Levin 2013). Others have focused on the 41 high-risk events, including seven catastrophic errors (Halsey 2013)—the result of a new categorization of events according to risk due to increased information contained in voluntary and electronic reports. FAA officials maintain that while the number of incidents reported has increased due to increased reporting, the skies have never been safer due to the agency's ability to analyze and learn from errors.

This report will examine the challenges and promise of developing and implementing incident reporting systems to evaluate and improve the performance of government agencies. Using a

case study of the FAA's ATO, this report will examine the agency's implementation of multiple incident reporting systems and derive a set of lessons learned for executives and managers considering the development of incident reporting systems.

Developing Performance Indicators

As Bob Behn declared in 2003, "Everyone is measuring performance" (Behn 586). The activity of managing and measuring the performance of public agencies has garnered the attention of public managers and scholars of public administration. The passage and implementation of the Government Performance and Results Act Modernization Act of 2010 (GPRAMA) provide agencies with the framework for performance management within the federal government.

Key Definitions

- **Performance Indicator:** a type of performance measure designed to measure the progress of the agency in achieving key agency outcomes.
- **Benchmarking:** the process of comparing an organization's performance indicators to those from other organizations.
- **Target:** specific goals for the organization to achieve by a specific date.

While the challenge of public and transparent agency performance reporting to ensure accountability is as old as the country itself (Ellig, McTigue, and Wray 2012), ensuring accountability has been made even more difficult as government activities increase in both scope and complexity. One of the most daunting challenges facing federal executives and managers is how to develop and use performance indicators that best illustrate the effectiveness of the organization in achieving its mission.

Many observers have noted that while agencies are collecting more information on their agency's processes, few are using that data to develop performance indicators or benchmarks to improve agency performance. Moynihan (2008; 2013) notes that the Government Accountability Office's (GAO) survey of government managers between 1997 and 2013 found that although agencies are collecting more data than ever before, there has been no significant change in the reported use of performance data for decision-making by managers (GAO 13-518).

A major driver of GPRAMA's passage was to address this deficiency. The Senate conference report on the GPRAMA notes that requirements for data-driven performance indicators and program reviews are "aimed at increasing the use of performance information to improve performance and results … Agencies are collecting a significant amount of information, but are not consistently using that information to improve their management and results" (U.S. Senate Committee on Homeland Security and Governmental Affairs, 2010, pg. 11).

The GPRAMA addresses the lack of performance information use by requiring agencies to develop a set of agency priority goals (APGs) and associated performance indicators that are reviewed on a quarterly basis. Additionally, Office of Management and Budget (OMB) Circular A-11 directs agencies to consider a variety of factors when developing indicators to measure performance in achieving APGs, including:

- Frequency
- Time
- Users and uses of the data
- Format of the data

- Method of data collection
- Context and analysis
- Costliness of collecting the data

(Source: 2012 OMB Circular A-11 Section 240)

The passage and implementation of the GPRAMA have led to further development of performance indicators across the federal government. However, the task of measuring agency performance through the use of data-driven indicators is often difficult for managers, given the "relative ambiguity of goals in public organizations and the nascent nature of performance-management efforts" (Nicholson-Crotty, Theobald, and Nicholson-Crotty, 2006). These authors point out that performance indicators come in many shapes and sizes and are often used by agency managers to accomplish a host of purposes other than strictly measuring agency performance. Kravchuk and Schack (1996) argue that no one measure or collection of measures is appropriate for all circumstances and that managers must settle on an "explicit measurement strategy" to develop effective performance indicators (p. 357). Behn (2003) argues that before choosing a performance measure, public managers first need to choose their purpose for the use of the indicator, including to evaluate, control, budget, motivate, promote, celebrate, learn, and improve (Table 1).

Table 1: Eight Purposes that Public Managers Have for Measuring Performance

Purpose	Performance indicator can help answer the following questions:
1. Evaluate	How well is my public agency performing?
2. Control	How can I ensure that my subordinates are doing the right thing?
3. Budget	On what programs, people, or projects should my agency spend the public's money?
4. Motivate	How can I motivate line staff, middle managers, nonprofit and for-profit collaborators, stakeholders, and citizens to do the things necessary to improve performance?
5. Promote	How can I convince political superiors, legislators, stakeholders, journalists, and citizens that my agency is doing a good job?
6. Celebrate	What accomplishments are worthy of the important organizational ritual of celebrating success?
7. Learn	Why is something working or not working?
8. Improve	What exactly should I do differently to improve performance?

Once managers have selected the purpose(s) that the indicator will measure, they then must examine the various characteristics of the indicator that must be understood to use it effectively. As Table 2 illustrates, the different purposes of indicators will tend to drive the type of data collected and used to construct the indicator. Berman and Wang (2000) and de Lancer and Holzer (2001) support this claim by finding that agencies typically divide performance indicators into three categories:

- Efficiency measures that assess costs and benefits, and outcome measures
- Programmatic impact and output indicators
- Policy and process components of the organization

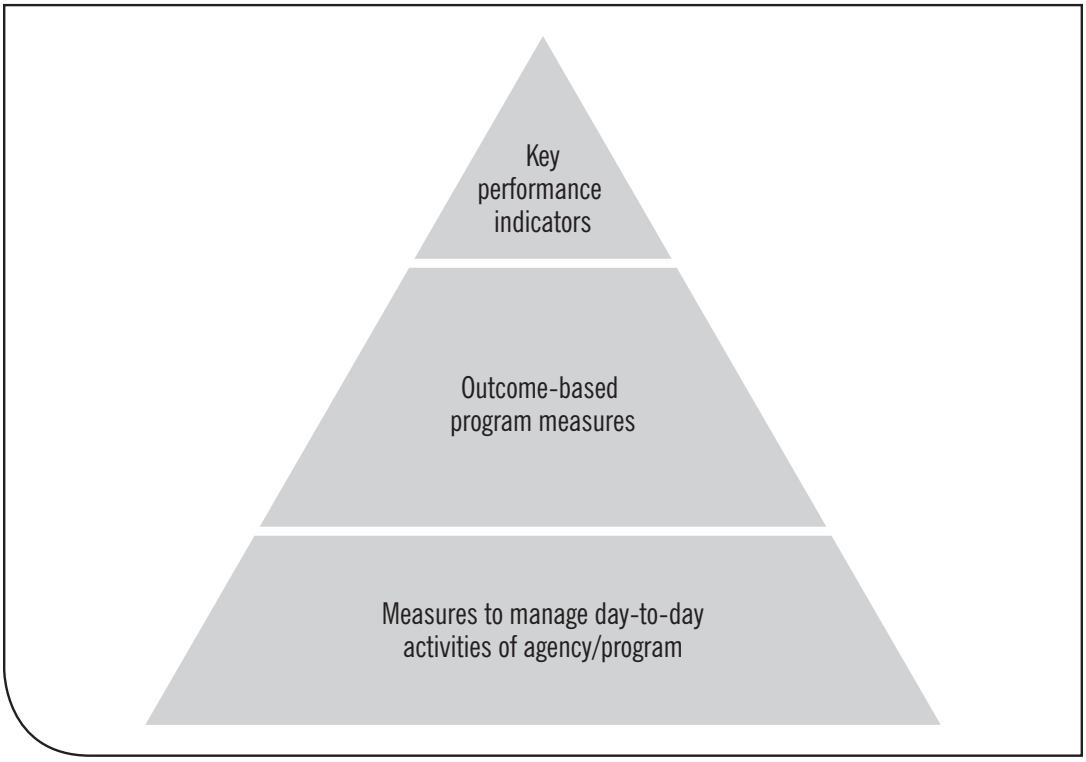
Table 2: Characteristics of Performance Indicator for Different Purposes

Purpose	Characteristic of Indicator
1. Control	Inputs that can be regulated
2. Budget	Efficiency measures (specifically outcomes or outputs divided by inputs)
3. Motivate	Almost-real-time outputs compared with production targets
4. Promote	Easily understood aspects of performance about which citizens really care
5. Celebrate	Periodic and significant performance targets that, when achieved, provide people with a real sense of personal and collective accomplishment
6. Learn	Disaggregated data that can reveal deviancies from the expected
7. Improve	Inside-the-black-box relationships that connect changes in operations to changes in outputs and outcomes

The different types of indicators used by agencies to measure performance are also graphically displayed in Figure 1:

- The base of the pyramid in Figure 1 describes output measures that are used to manage the day-to-day performance of employees within the agency.
- Above the output indicators are outcome or programmatic indicators designed to measure the overall impact of the program on society.
- Key performance indicators are a subset of the outcome-based measures that are used to communicate the performance of the agency to external stakeholders. Often, these key performance measures can be iterations or extensions of the outcome-based program measures used to internally manage agency performance.

Figure 1: Hierarchy of Performance Indicators



Source: State of Indiana Office of Management and Budget

Challenge of Selecting Performance Indicators

While the above process of indicator selection and implementation appears logical and linear, many challenges are often presented to managers as they identify data and indicators to measure agency performance. Nicholson-Crotty, Theobald, and Nicholson-Crotty (2006) note that even when managers agree on the purpose of an indicator, there may be multiple plausible measurements of the concept being measured. Additionally, these managers may face internal or external pressure to select a particular method or indicator to measure outcome, output, or even input. Internally, managers may face pressure from different offices or units to adapt or implement an indicator that does not suggest performance improvement when compared with another internal branch of the agency.

Additionally, external entities including Congress, the OMB, stakeholders, and the media can impose pressure on agencies to adopt and implement a series of performance indicators that measure a highly observable aspect of agency performance. Behn (2003) notes that often managers of a public agency do not have the freedom to choose their own performance measures, and that external parties such as Congress and OMB may require the agency to report a series of indicators that the agency may have little control over. While agencies must report the indicators requested by these external actors, they also have an obligation to create and implement a series of performance indicators that they will use to manage agency performance. Additionally, as Behn (2003) notes, agencies can use the indicators imposed on them by external actors to supplement their existing measures or to achieve other purposes, such as motivating employees.

Another challenge facing public managers in developing indicators is the level of observability of the agency outputs and outcomes used to construct indicators. How easily the work of the agency can be quantified and explained to external entities as an indicator affects the level of pressure, and therefore autonomy, that an agency will have in achieving its mission.

In his seminal work, *Bureaucracy*, James Q. Wilson (1989) developed a typology (Appendix I) of agency types based on the observability of both outputs (what agencies do on a day-to-day basis) and outcomes (how the work of the agency changes the state of the country). Significantly, Wilson argues that the level of observability of the outcomes and outputs of an agency determine what tasks the agency will complete. As Peters and Waterman (1982) note, "What gets measured gets done." However, as Behn (2003) counters, "what people measure often is not precisely what they want done" (p. 599).

A focus on the observability of outputs and outcomes that are performance indicators highlights the importance of internal agency performance reporting and of data collection systems that gather information on how both frontline employees and managers are performing. Many executives and managers across agencies have spent valuable resources to collect and analyze data on how effectively employees are implementing the mission of their organizations. The focus on the collection of data on employee performance is particularly important in high-reliability agencies, where poor performance can result in high-impact consequences such as a loss of life or large-scale disasters.

The Emergence of Incident Reporting Systems

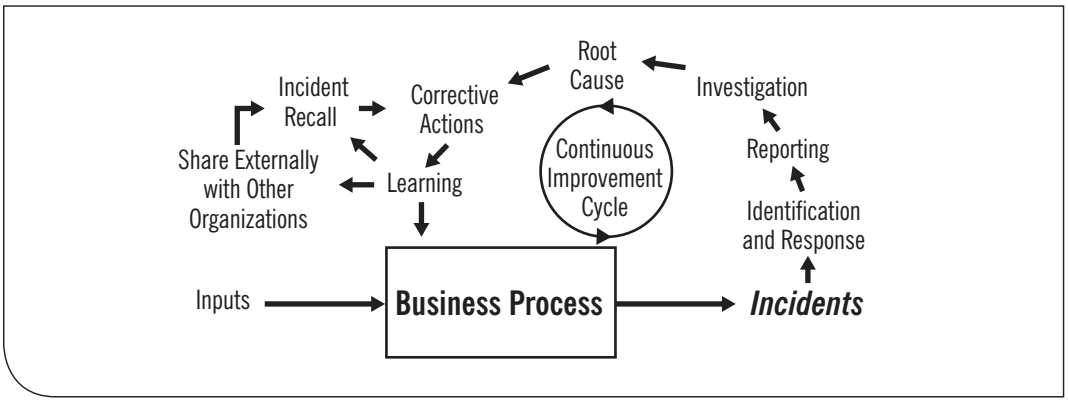
As many functions of both private industry and operational government agencies become increasingly complex and interconnected, the risk for serious accidents and disasters increases. When the importance of avoiding serious accidents or incidents is key to the work of the organization, it is commonly referred to as a high-reliability organization. Managers of high-reliability organizations often try to prevent what are known as normal accidents, those which are a normal

consequence of the close coupling and interconnectedness of a large number of complex variables (Sagan 1993). There are often several smaller incidents that can serve as warning signs or learning opportunities to prevent accidents (Turner 1978; Cooke and Rohleder 2006).

Incident reporting systems are a common tool used by high-reliability organizations to prevent accidents. Employees report, or equipment detects, incidents or errors committed by employees. These are then analyzed for cause and used to develop corrective actions. Significantly, while incidents may be viewed as unwanted outputs of an agency or industry process, they are a vital piece of performance information that, if used properly, can foster continued organizational improvement.

Figure 2 illustrates a typical incident reporting system that is often used by high-reliability organizations. The process begins with the identification and response to incidents by frontline employees, who then either voluntarily report an incident or are detected committing an incident. Following receipt of the incident report, a team of managers and employees typically investigate the incident and identify its root cause. Through this investigation of the root cause, the team of managers and employees may learn of similar incidents occurring in other parts of the organization that may have the same root cause and can be mitigated though a corrective action. Once the agency or business implements a corrective action, it then must use a quality control process to investigate (or learn) if the corrective action solved the causal problem identified through the incident reports. Finally, agencies or companies then share the information regarding the success of the corrective action externally with other organizations.

Figure 2: Incident Reporting System



Source: Adapted from Cooke and Rohleder 2006

There are several examples of high-reliability agencies that use incident reporting systems to help mitigate the risk of serious accidents or disasters. The FAA uses voluntary incident reporting systems in partnership with U.S. air carriers to identify risks and hazards both at the air carrier and within the FAA (Mills 2010). The Food and Drug Administration (FDA) requires firms that use and manufacture medical devices to report device-related injuries or deaths to the agency.

The Federal Emergency Management Agency (FEMA) operates the National Fire Incident Reporting System to identify national trends and best practices in firefighting. Recently, a *Washington Post* investigative report found that the National Security Agency (NSA) used an incident reporting system to identify compliance incidents (overreach on the part of analysts, improper entry to data, etc.) during its surveillance activities (Gellman 2013). Also, a USDA

pilot program (called the Hazard Analysis and Critical Control Point-Based Inspection Models Project, or HIMP), designed to encourage more inspections and reporting of hazards in the oversight of pork processing plants, has come under fire from Congress and the GAO as the number of violations at those plants have increased under the pilot program (Kindy 2013). Finally, the medical community relies on incident reporting to identify medication administration errors (MAE) at both the facility and at the national level through the MEDMARX reporting system (Wakefield 2001).

While the use and success of incident reporting systems have increased dramatically over the past decade, many managers in high-reliability agencies have faced challenges when trying to use reported incidents as indicators of an agency's commitment to performance improvement. Many of these behavioral challenges are related to the system design challenges associated with incident reporting systems: ambiguity about incident causation, the politicized environments in which incident investigation takes place, the human tendency to cover up mistakes, and the tendency toward secrecy both within and between competing organizations (Sagan 1993; Cooke and Rohleder 2006). The next section of the report explores the challenges of developing performance indicators with information from incident reporting systems in the FAA ATO.

Case Study: The Evolution of Safety Performance Indicators in the Federal Aviation Administration's Air Traffic Organization (ATO)

Background on the FAA and the ATO

The Federal Aviation Administration (FAA) is the primary regulatory agency in charge of air transportation in the United States. In addition to providing safety oversight of air carriers, airmen, and aircraft operating in the United States, the FAA is also responsible for serving as the operator of the nation's air traffic control system. Specifically, the FAA's Air Traffic Organization (ATO) is responsible for ensuring aircraft are safely separated in the National Airspace System (NAS) through the operation of 21 (20 domestic; one international) En Route Air Traffic Control Centers, 166 terminal radar control facilities (TRACONs), and another 380 air traffic control towers operated by the FAA or contract controllers. The FAA employs over 16,000 air traffic controllers and managers to operate these facilities, many of whom are represented by the National Air Traffic Control Association (NATCA).

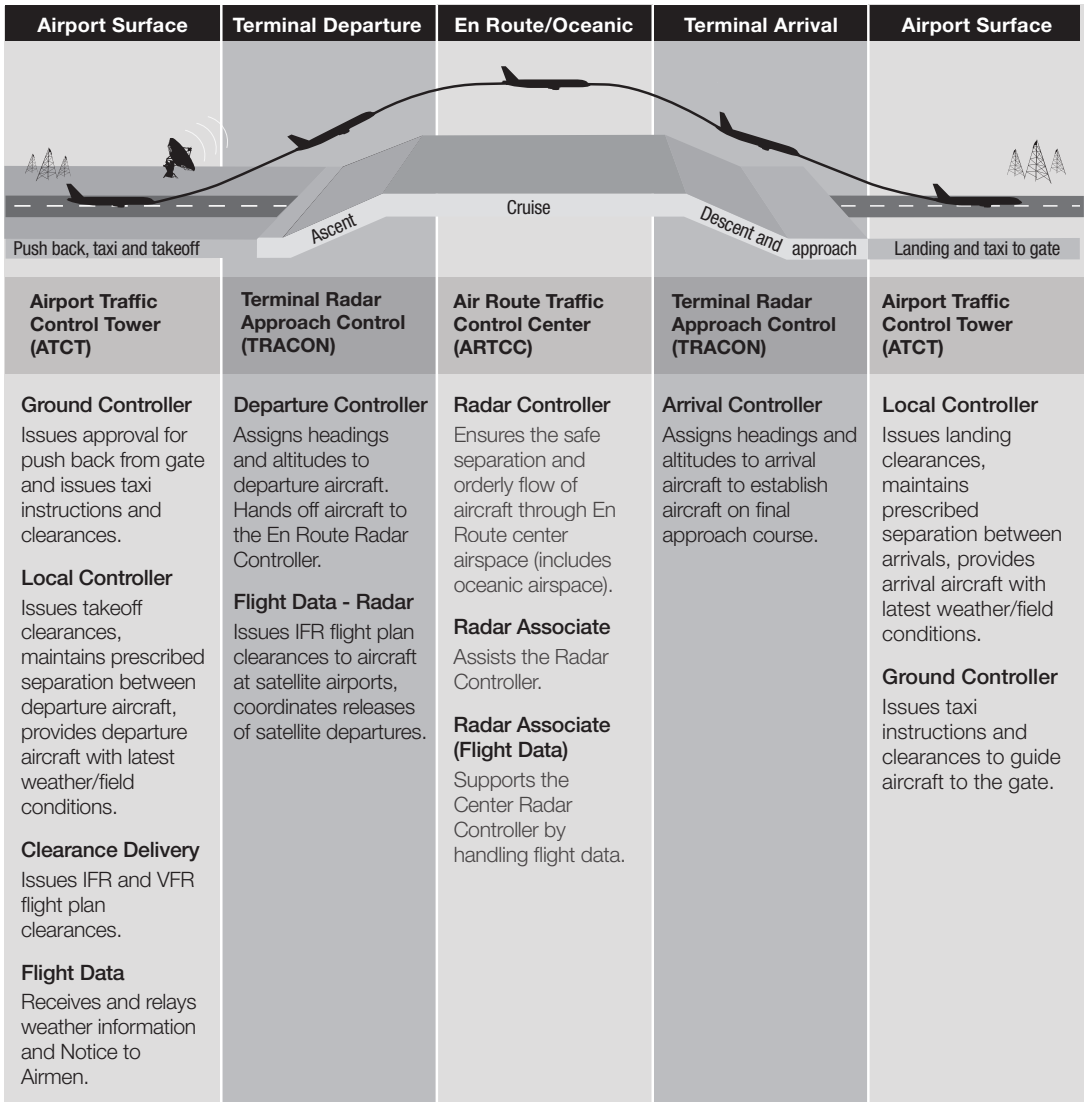
Figure 3 provides an overview of the levels of flight and the division of labor within the ATO. Air traffic controllers in airport towers issue takeoff clearances and maintain separation standards between arriving and departing aircraft, and manage the sequencing of departing and arriving aircraft. TRACON controllers assign headings and altitudes to departing and arriving aircraft (typically within 40 miles of the airport) while also "handing" traffic to en route controllers. En Route Center controllers separate aircraft above flight level 170 (17,000 feet) while also providing advisory services to smaller airports without radar coverage.

An Introduction to Incident Reporting at the FAA

At any given time, there are over 7,000 aircraft flying over the United States. To maintain safe distances between them, the FAA established minimum separation standards based on the aircraft's phase of flight and size of aircraft. These separation standards help to reduce the risk of serious incidents, including midair collisions and runway incursions. Air traffic controllers in TRACONs and air traffic control towers are responsible for maintaining horizontal separation of three miles and vertical separation of 1,000 feet. In the En Route environment (above 17,000 feet), center controllers provide communications and instructions to pilots to maintain horizontal separation of five miles (higher separation standard because of higher speed) and vertical separation of 1,000 to 2,000 feet (Figure 4).

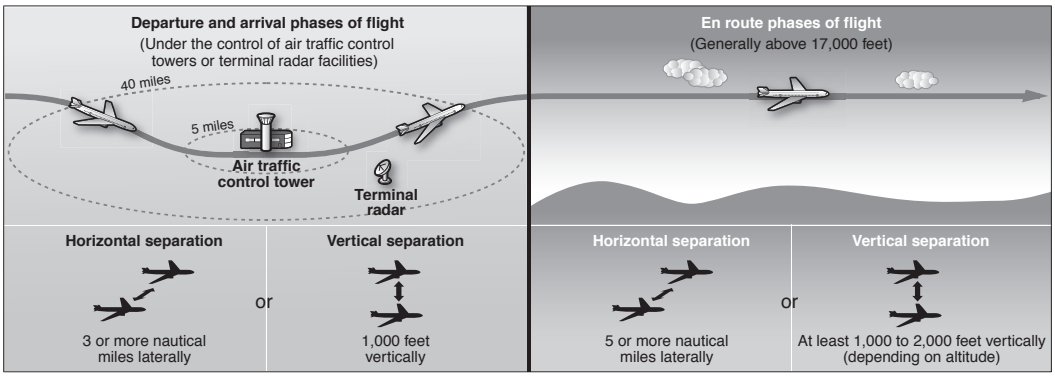
In the past, the FAA measured the performance of air traffic controllers in maintaining safe separation standards through the rate of operational errors (OEs) reported by air traffic control facilities. When a loss of standard separation (LoSS) occurred, those incidents were investigated to determine whether the LoSS was the result of a pilot deviation, an operational error (OE), or an error as the result of a misguided policy or equipment failure. LoSS was reported and investigated by air traffic personnel in the tower, TRACON, or En Route Center where the incident occurred. The controller and his supervisor would complete a 25-page report on the incident and would send that report to the ATO service center and eventually to ATO headquarters in Washington (FAA ATO Interview B June 4, 2013).

Figure 3: Air Traffic Control During the Stages of Flight



Source: FAA Controller Workforce Plan 2012

Figure 4: FAA Separation Standards in Terminal and En Route Environments



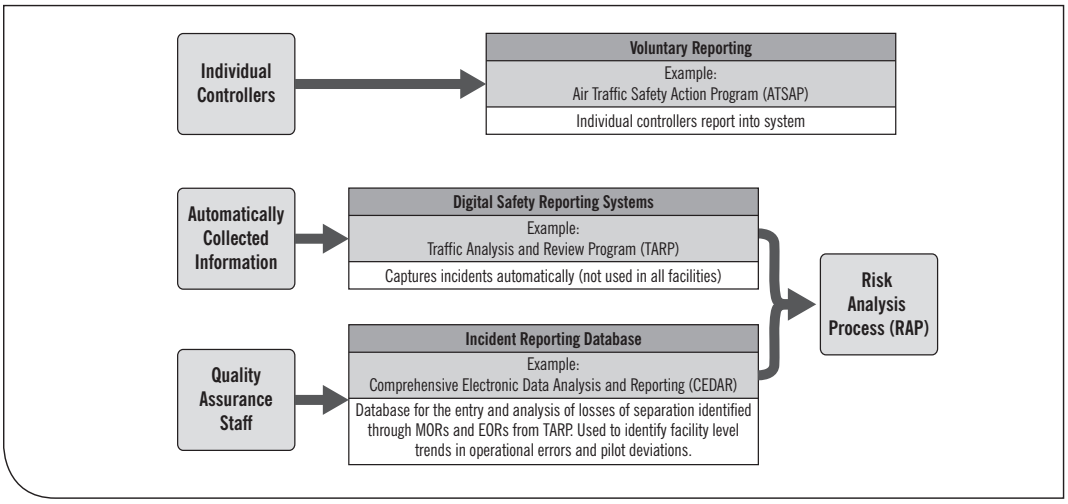
Source: GAO

Table 3: Components of FAA ATO Incident Reporting System

Type of Incidents	Description
Loss of Standard Separation (LoSS)	An incident that occurs when two aircraft fail to maintain horizontal separation of three miles and vertical separation of 1,000 feet in the TRACON environment, or horizontal separation of five miles (higher separation standard because of higher speed) and vertical separation of 1,000 to 2,000 feet in the En Route environment.
Operational Error (OE) (1989–2011)	An incident that occurs when controllers fail to apply or follow the procedures that enforce separation and allow aircraft to end up too close to each other or to an obstruction.
Pilot Deviation (PD) (1989–Present)	An incident that occurs when the actions of a pilot violate a Federal Aviation Regulation, including failing to maintain safe separation standards.
Individual Reporting Systems	
Mandatory Occurrence Reporting (MOR) (1989–Present)	An occurrence involving air traffic services for which the collection of associated safety-related data and conditions is mandatory, including a loss of standard separation on final approach.
Air Traffic Safety Action Program (ATSAP) (2008–Present)	A voluntary reporting program for air traffic control personnel that allows them to enter a qualitative self-assessment of an incident.
Automated Reporting Systems	
Electronic Occurrence Reporting (EOR) (2012–Present)	An alert identified by an automated system, such as the Traffic Analysis and Review Program (TARP) that automatically uploads into the Comprehensive Electronic Data Analysis and Reporting (CEDAR) tool
Terminal Area Reporting Program (TARP) (2012–Present)	An automated software system used to detect and quantitatively measure losses of separation at air traffic terminal facilities.
Incident Analysis Tools and Processes	
Comprehensive Electronic Data Analysis and Reporting (CEDAR) (2010–Present)*	Database for the entry and analysis of losses of separation identified through MORs and EORs from TARP. Used to identify facility level trends in operational errors and pilot deviations.
Risk Analysis Event (RAE) (2012–Present)	Losses of standard separation occurrences that have a measure of compliance of less than 66%.
Risk Analysis Process (RAP) (2012–Present)	The process used to identify key elements in RAEs and to give each RAE a value. ATO Safety uses occurrence data for input into a risk analysis tool which will then provide a risk assessment value associated with an individual occurrence. The process is conducted by a dedicated group of persons with air traffic control (ATC) experience and air transport-rated pilots with commercial experience.
Incident Severity Indicator	
System Risk Event Rate (SRER) (2011–Present)	The rate of the most serious loss of standard separation for every thousand instances of loss of standard separation within the NAS. The SRER metric will first be calculated using only aircraft-to-aircraft loss of standard radar separation.

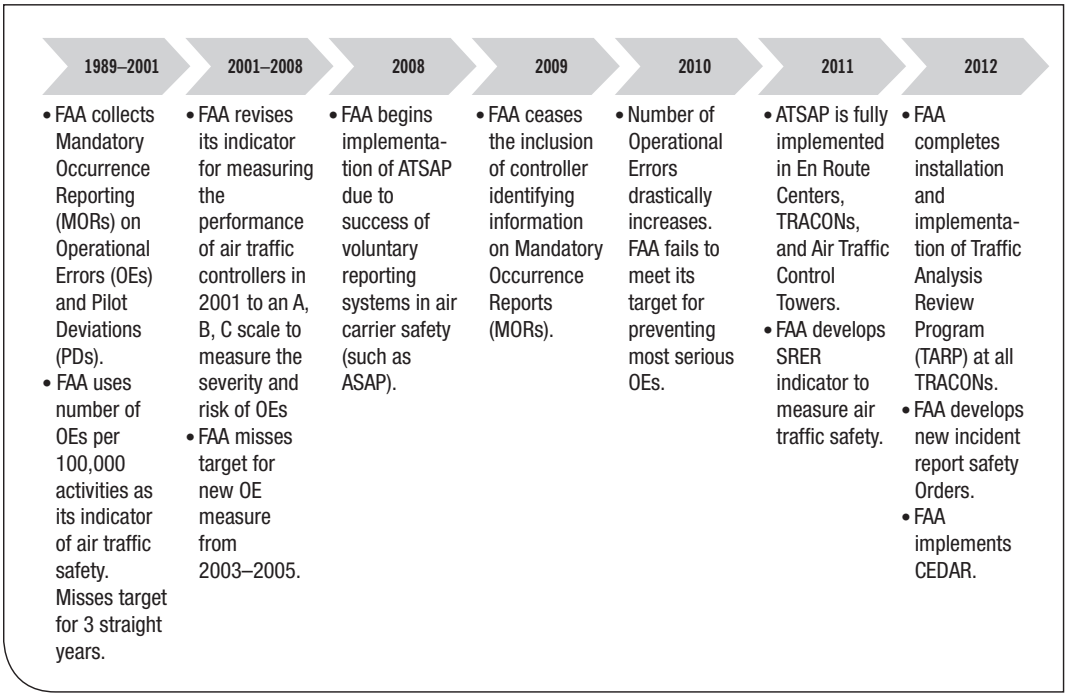
** Formerly known as Air Traffic Quality Assurance (ATQA)*

Figure 5: The Compartmentalized Nature of ATO Incident Reporting



Source: GAO

Figure 6: Timeline of Evolution of FAA Incident Reporting Systems



Tracking Operational Errors

From 1989 to 2001, the FAA measured both facility-level and agency-level performance in the area of air traffic control safety through the number of OEs (again, a subset of all LoSS) reported by facilities (FAA Performance and Accountability Reports, 1999–2002). Specifically, the agency would set baselines and targets for each facility and for the agency as a whole that would determine the level of performance at each. Significantly, under the union contract with NATCA, managers' pay raises were often tied to meeting or exceeding facility level performance metrics, including the number of OEs. This created a negative incentive to report OEs:

> So I mean it is human nature and it is understandable that if you're told you can have no more than 10 operational errors in a year and you're at nine and next year's pay depends on that, you're probably not going to have 10. You're probably going to make sure that nobody reports that 10th one. So that was not a good incentive plan to keep operational errors low, because it didn't really reflect the total number of operational error incidents that were occurring (FAA ATO Interview A, May 16th, 2013).

Despite the incentive not to report, the FAA continually missed its targets for agency-wide OEs.

Prior to 2001, the FAA used the number of total OEs per 100,000 activities as its metric for air traffic control safety. However, from 1998 to 2000, the agency missed its targets for OEs.

Year	OE Target (OEs Per 100,000 Activities)	Actual OEs Per 100,000 Activities
1998	0.504	0.55
1999	0.496	0.57
2000	0.486	0.68

Source: FAA Performance and Accountability Reports (1997–2001)

In 2001, following three consecutive years without meeting the OE target, the agency revised its indicator for measuring the performance of air traffic controllers in maintaining safe separation standards. The agency implemented a severity index that classified all OEs according to four levels (Figure 7):

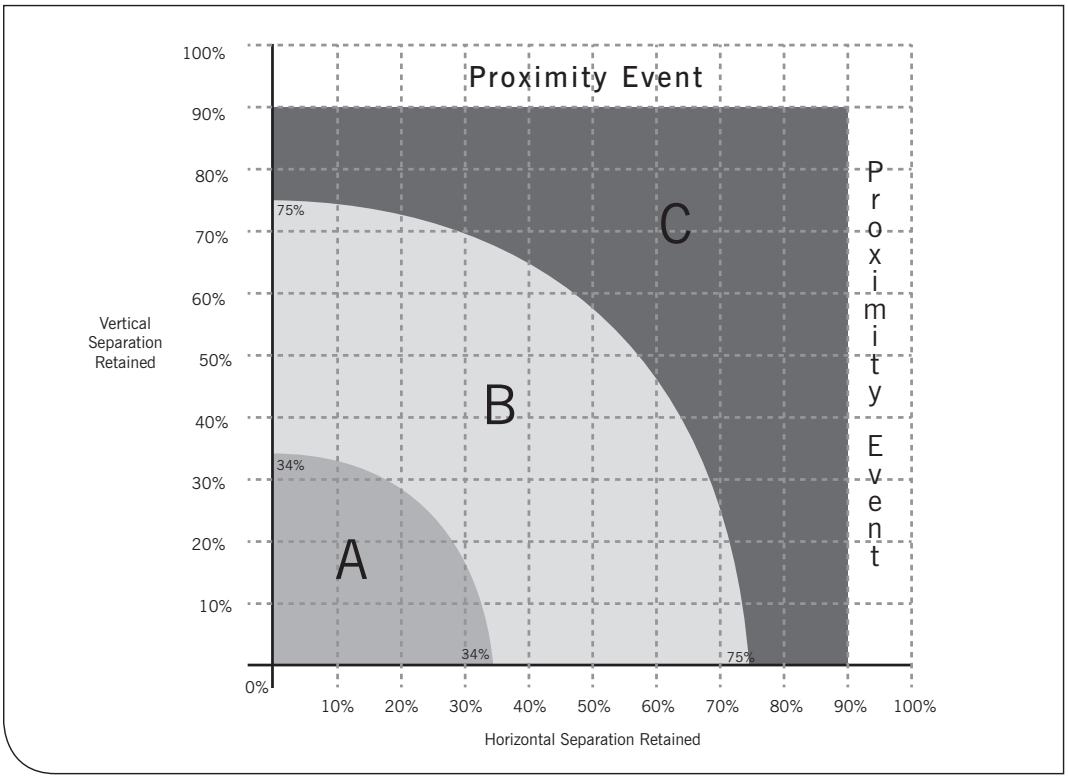
- Proximity Events: An OE where 90% or greater separation is retained in either the horizontal or vertical plane and does not include any violation of wake turbulence separation minima. Although proximity events are technically OEs, they are not viewed as a threat to aviation safety.
- Category A (high risk): This is the most severe type of OE and involves a loss of airborne separation where the separation conformance is less than 34%.
- Category B (moderate risk): A loss of airborne separation where the separation conformance is greater than or equal to 34% but less than 75%.
- Category C (low risk): A loss of airborne separation where the separation conformance is greater than or equal to 75%, but the horizontal and vertical separation retained is less than 90% of the required separation.

Using a strategy similar to measuring the agency's performance on runway incursions, the agency would assess the performance of controllers in preventing LoSS by the number of Category A and B OEs on an annual basis. Significantly, to assess the severity of the loss of separation, the FAA was primarily relying on the report submitted by the controller who witnessed the incident.

The DOT inspector general (DOT-IG) had previously raised concerns that incidents were being unreported in the TRACON and tower environments due to a reliance on self-reporting by controllers. Additionally, the DOT-IG was concerned that air traffic control facilities retained voice tapes and radar data (that could be used to find unreported incidents or confirm the severity of reported OEs) for only 15 days (DOT-IG Reports AV-2001-11, 2000; AV-2004-85, 2004).

Thus, because of the reliance on self-reporting of incidents and the pressure to meet performance targets for the number of OEs, the move to a severity classification of OEs did little to

Figure 7: FAA Loss of Separation Categorization Tool (2001–2011)



Source: FAA Order 7210-56C Appendix 9

improve the reliability of the performance data being collected. As one ATO official notes, the controllers' reports tended to be biased in a direction that would make their mistake's severity look less than it may have been (FAA ATO Interview B, June 4, 2013).

Implementation of Voluntary Incident Reporting (ATSAP)

In 2005, the FAA began to shift its safety oversight of the ATO toward a more risk-based (rather than punitively based) paradigm, largely due to the success of voluntary reporting programs used by the agency to oversee the safety of air carriers (Mills 2010). The goal of a risk-based safety paradigm is to encourage the reporting of as many incidents as possible to gain an understanding of the trend of incident reporting, along with the causal factors behind the incident.

Understanding the causal factors of an incident allows analysts and senior agency officials to design corrective actions to remediate the underlying causes. To foster a free exchange of information between controllers and their supervisors, the FAA created the Air Traffic Aviation Safety Program (ATSAP) in 2008. Largely modeled after the Aviation Safety Action Program (ASAP) used by commercial air carriers, ATSAP allows controllers to voluntarily disclose incidents to an Event Review Committee (ERC) comprised of representatives from ATO, the FAA Office of Air Traffic Safety Oversight Office (AOV), and the controllers union, NATCA (FAA ATO Interview A, May 16, 2013).

To incentivize the open reporting of errors, ATSAP guarantees that those who report incidents will be provided anonymity and protected from disciplinary action (FAA Order 7200.20). Once an ATSAP report is received, the ERC reviews the report, determines any necessary corrective

action (training, procedure changes, etc.), determines the risk and severity of the incident, and submits the report. ATSAP ERCs in each service area meet twice a week to review and validate each narrative ATSAP report. Once the report is submitted, it is hosted on a local server provided through a contractor. The aggregated de-identified data may also be analyzed at the facility level to identify trends and possible corrective actions through a program called Partnership for Safety. Also, anonymous ATSAP data is shared with a number of air carrier ATSAP programs through the Confidential Information Sharing Program (CISP), and with Aviation Safety Information Analysis and Sharing (ASIAS) operated in partnership with the MITRE Corporation (FAA ATO Interview A, May 16, 2013).

The ATSAP program was fully implemented in October of 2010 and has collected over 70,000 reports since its inception. It is important to note that while ATSAP is a valuable source of data on the performance of frontline employees, the incidents contained in ATSAP reports may not all be part of the agency's count of OEs reported to Congress or the public. Because of the confidential nature of the ATSAP data and the fear that facility managers would take action against individual controllers, the FAA agreed to not merge ATSAP data with its other sources of air traffic control safety data (FAA ATO Interview B, June 4, 2013). Both the DOT-IG (Report AV 2012-152) and the GAO (Report 12-660T) have raised concerns that by not integrating ATSAP reports into the FAA's other sources of air traffic control safety data, the FAA may be underestimating the number of serious OEs taking place in the national airspace system.

Expanding Incident Reporting Through Technology (TARP and CEDAR)

At the same time the agency was implementing ATSAP, the ATO was taking steps to also improve the robustness and analytical capability of its traditional incident reporting systems. Following a DOT-IG report in 2008 that found intentional and widespread misreporting of OEs by controllers at the Dallas-Fort Worth TRACON, the FAA began work on developing and implementing an electronic incident reporting system. The Traffic Analysis Review Program (TARP) is an automated electronic incident detection system that records LoSS in TRACONs and reports them to an electronic database called Comprehensive Electronic Data Analysis and Reporting (CEDAR).[1] The electronic occurrence reports (EORs) generated through TARP and matched with mandatory occurrence reports (MORs) from facility managers and employees provide a rich source of operational performance data. The FAA completed implementation of TARP at all of its TRACONs in 2012 and is currently working on implementing TARP in the En Route environment.

The increase in data collection on the number and surrounding context of incidents through TARP led to a drastic increase in the number of OEs reported in 2010, its first full year of operations. Also, although ATSAP reports are not considered in the official count of OEs, the FAA's commitment to non-punitive reporting led to more incident reporting by frontline managers and controllers through traditional reporting methods. To encourage more reporting from frontline employees that would be captured in CEDAR, the FAA agreed to remove identifying information from reports in 2009. As one ATO official notes:

> As you decriminalize reporting, you have gotten a lot more reporting. So, between 2009 and 2010 the reporting of operational errors just skyrocketed. I mean that was the year, and it has been pretty constant since 2010. The reason was not just ATSAP,

1. The En Route environment has a similar system called the Operational Error Detection Program.

> because we just don't report what's only in ATSAP, but it was the decriminalization of reporting. So, I can now raise my hand and go write an ATSAP report or file a MOR and not be disciplined (FAA ATO Interview A, May 16, 2013).

The benefit to the FAA of increased reporting through TARP, MORs, and ATSAP was that it allowed for better quality assurance and quality control processes to improve the performance of its controllers and the safety of air traffic in the United States. However, as the GAO and DOT-IG have both noted, removing the identifying information from MORs also made it more difficult for frontline managers to evaluate and help to improve the performance of individual controllers (GAO 12-660T; DOT-IG AV 2013-046).

Internally, the FAA believed that its shift to a risk-based, non-punitive safety oversight system was resulting in increased safety. Yet many outside stakeholders, including members of Congressional authorizing and appropriations committees and the media, began to question why the FAA was exceeding its target levels for OEs. Specifically, between FY 2008 and FY 2011, the FAA did not meet its target goals for reducing the number of the most severe categories of OEs (DOT IG Report AV-2013-046). This resulted in several GAO, DOT-IG reports, Congressional hearings, and media stories on the increase in the number of OEs. Much of the coverage focused on why controllers were committing more errors, as the volume of air traffic was remaining fairly steady (Halsey 2012). Behind the scenes, FAA officials were trying to explain to stakeholders that the increase in the number of OEs was not an indication of an actual increase in errors, but rather an increase in incident reporting:

> We haven't been able to convince the IG, GAO, and the Hill that removing the controller's name from an OE report, the introduction of ATSAP and stopping disciplinary action after reporting OEs is what caused the reporting of OEs to skyrocket (FAA ATO Interview A, May 16, 2013).

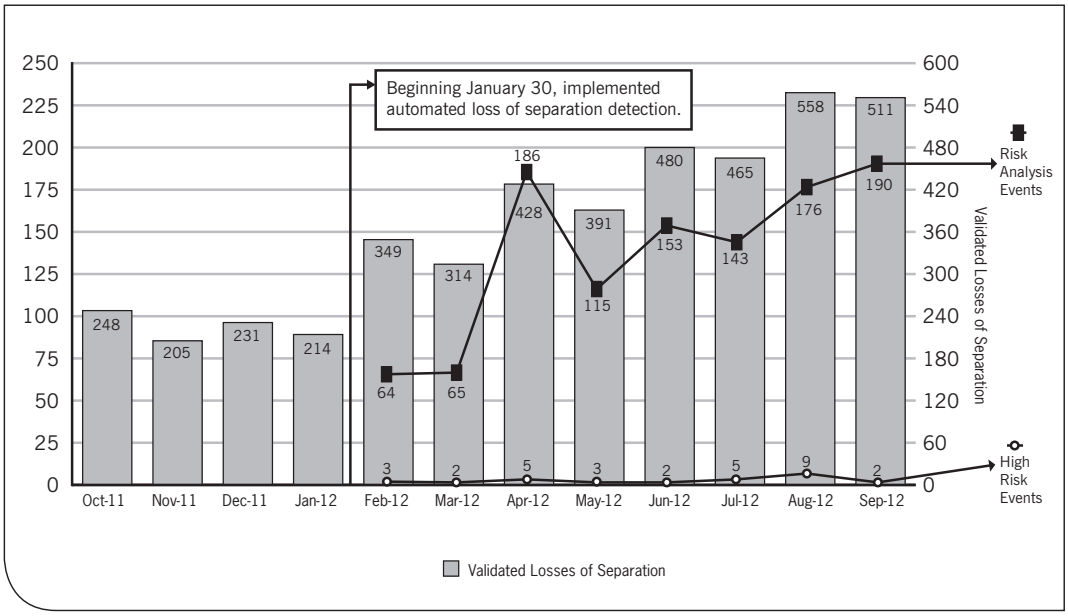
Developing a New Incident Reporting Metric: The System Risk Event Rate (SRER)

In 2011, FAA officials decided to develop a new indicator that would provide a better view of the agency's performance and capture the risk-based nature of their safety oversight system. As one official notes, the consensus among agency leadership regarding the indicator measuring severe OEs contained within the strategic plan was, "We have to fix the problem, we have to change it." This is because the former metric of OEs was not moving the needle on safety (Former FAA Official Interview, May 16, 2013). As the agency was developing the indicators that would eventually become part of its new strategic plan, Destination 2025, there was considerable debate internally and between DOT and OMB over which indicator would replace the rate of OEs. One proposal was to look at average amount of separation maintained through all incidents (FAA ATO Interview B, June 4, 2013). However, the FAA, DOT, and OMB decided on measuring and reporting the ratio rate of the most risky losses of separation compared to all losses of separation. This new indicator, called the System Risk Event Rate (SRER), would allow the FAA to control for the increased level of reporting of events by presenting the indicator as a rate of high-risk losses of standard separation to all such losses of separation. Therefore, as the overall number of separation losses reported (the denominator) grew, the SRER rate would be driven down because the electronic reporting system, TARP, was capturing an increasing number of less serious losses of separation (FAA ATO Interview B, June 4, 2013). The increased level of reporting of losses of separation as a result of the implementation of TARP is illustrated in Figure 9.

Figure 8: System Risk Event Rate Formula

$$\text{SRER} = \frac{\sum(\text{Serious Loss})}{(\text{Number of LoSS Events Identified Through CEDAR})} \times 1000$$

Figure 9: Increases in Identification of LoSS and Risk Analysis Events due to TARP Implementation (2011–2012)



Source: FAA 2013

The new SRER metric required the FAA to perform a higher level of analysis on its incident reports than it had in the past. In early 2012, the FAA released a series of new quality assurance and quality control orders outlining how the agency would categorize the risk of LoSS reported into CEDAR through TARP and through MORs. Once a loss of separation is identified through TARP and reported to CEDAR, those events with less than 66% of required separation will become known as Risk Analysis Events (RAEs) and will be subject to further investigation and inquiry by a team of ATO analysts at each service area (FAA Order 7210.633). Each group of analysts, known as Risk Analysis Process (RAP) panels, is composed of ATO officials and former commercial pilots who validate the severity of the loss of separation, determine the closing rate of the aircraft, and determine the probability of incident repeatability by examining the state of the system at the time of the incident (FAA ATO Interview B, June 4, 2013).

RAP panels in each service area meet three to four times a week to examine and validate the highest-risk incidents reported to CEDAR through electronic TARP detection, using a risk and severity matrix. This frequent validation and assessment of the data allows for continual updates on ATO's safety performance. The RAP panels come to a consensus on the overall level of risk and severity associated with the RAE and assign it a value. Figure 10 illustrates the risk analysis matrix and the categorization of all risk analysis events in FY 2012. The highest categorization of RAEs identified through TARP and CEDAR becomes the numerator of the SRER metric.

Figure 10: ATO Risk Analysis Matrix with FY 2012 RAEs

SEVERITY / LIKELIHOOD	Minimal 1	Minor 2	Major 3	Hazardous 4	Catastrophic 5
Frequent 5	0	0	1	1	0
Probable 4	1	7	10	3	4
Remote 3	4	78	108	19	2
Extremely Remote 2	100	427	430	67	1
Extremely Improbable 1	4	0	4	0	0

HIGH RISK	41
MEDIUM RISK	182
LOW RISK	1048
Total RAE	1271

- **PROXIMITY**
- **CLOSURE RATE**
- **REPEATABILITY**
- **SEVERITY**

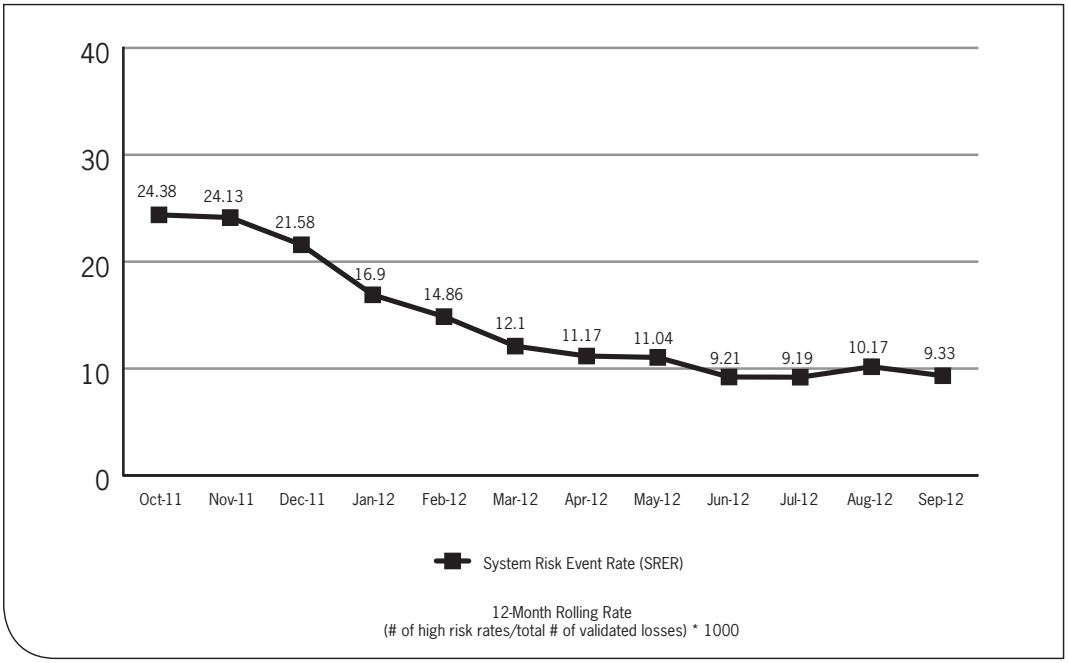
Source: FAA 2013

An interesting aspect of RAP and the new process is that the ATO is actually reviewing and investigating fewer losses of separation than it had in the past (GAO 12-24).The new FAA safety orders moved the ATO's standard for investigating LoSS from less than 90% separation standard to less than 66%. FAA officials noted that this transition allowed the agency to maximize its limited resources while also focusing its analytical capabilities on the highest risk events (FAA ATO Interview B, June 4, 2013).

Since SRER's inception as the primary indicator for measuring the overall safety of the air traffic system, the FAA has achieved its SRER target of no more than 20 serious losses of separation for every 1,000 LoSS for most months in FY 2012 (Figure 11). ATO officials attribute this success to increased electronic reporting from TARP (ATO Interview B, June 4, 2013). However, the GAO and DOT-IG remain concerned that the FAA has not fully integrated all sources of safety data, which limits its ability to identify potential risks and take corrective actions to improve performance (GAO 12-660T; DOT-IG AV 2013-046). The data from ATSAP and from the TARP and CEDAR reporting systems remain somewhat compartmentalized, primarily for political rather than technical reasons. However, the agency has taken steps to merge these three sources of data.

Recently, the FAA included a new indicator in its core compensation indicators that identifies the percentage of the top five risk categories identified through the ATSAP and RAP reporting systems that the agency mitigates in a year. The new top five risk categories indicator allows the FAA to show meaningful action on the areas of highest risk identified through its collection of information on losses of separation. However, because there are only five risk categories identified each year, FAA officials admit that there is a tendency to select short-term mitigations that can realistically be accomplished in one year, rather than more ambitious and complex mitigations. (FAA ATO Interview B, June 4, 2013).

Figure 11: System Risk Event Rate (SRER) FY 2012 Performance



** FAA Target is 20 high risk RAEs/Total # of Validated Losses*

Source: FAA 2013

Top Five Risk Categories Indicator

In 2011, the FAA began tracking and identifying the most risky incidents that were reported through both its individual and electronic reporting systems. The risk analysis process (RAP) panels identify the level of risk and severity associated with incidents determined electronically through TARP, while ATSAP review panels assess the risk and severity associated with voluntary reports through ATSAP. In 2012, the ATO identified the tendency of landing aircraft to come close together as they were turning into their final approach to airports with closely spaced parallel runways. Because of the information contained in both EORs (through TARP) and in voluntary reports through ATSAP, the FAA implemented new rules and procedures to instruct pilots to make a tighter turn in final approach to reduce risk of a potential collision.

Top Five Risks Identified Through RAP and ATSAP (2012)

- Turning planes for their final runway approach
- Planes coming too close when landing on parallel runways
- Enforcing "go-around procedures" (maneuvers when planes are told to abort their final approach)
- Proper guidance to pilots who don't hold their designated altitudes
- Coordination between controllers as they pass responsibility for an airplane from one controller to another

The Influence of External Stakeholders in Designing Agency Indicators Using Incident Reporting

One significant facet of FAA performance management is the interaction and occasional tension between the agency and DOT in crafting its own strategic plans, performance indicators, and priority goals. The FAA has long been considered an exemplar of how agencies should implement performance management in the government (Center for American Progress 2011; Partnership for Public Service 2011). Therefore, when the GPRAMA was passed, planners within the FAA were already engaged in many of the activities prescribed by the new law, including monthly internal meetings to discuss performance and quarterly meetings with DOT to review performance indicators (FAA Planning Official Interview, May 13, 2013). Several FAA officials note that the agency's quarterly meetings with DOT had been integrated with another parallel meeting discussing regulatory or legislative items on the agenda. One former FAA official notes that the first quarterly meetings under the Modernization Act were only partially focused on performance management:

> GPRAMA required DOT (not FAA) to hold quarterly meetings on performance, so DOT was meeting with all of the modes individually, including FAA. But ATO already had a standing weekly meeting with the DOT deputy secretary on their performance, so this quarterly check-in was a bit redundant. There were times DOT would ask me for more information on our quarterly report, and I would go back to AVS (Aviation Safety) or ATO and ask for more information as necessary. But as far as GPRA goes, DOT was already far out ahead of the quarterly requirement (Former FAA Official Interview, May 16, 2013).

However, FAA officials note that the quarterly meetings with DOT have become more focused on performance, particularly in evaluating and crafting priority goals for the upcoming year (FAA Official Interview A, May 16, 2013).

One new challenge for the FAA was how to work with DOT and OMB to select the subset of goals and indicators from Destination 2025 and the core compensation goals that would become part of the DOT's priority goals. A senior FAA official describes the process of how the agency internally prioritizes goals and indicators that are suggested to DOT and to OMB to be made public on Performance.gov:

> We do have sort of a hierarchy of a few … items that we report to the department, a fewer number … that go to OMB. And then a broader range that are centrally looked at here, at FAA and then an even richer set that are looked at [by] lower levels in the different organizations for determining how they're doing, and I think that's entirely appropriate. You don't want one of those metrics the subject of … having to explain to a well-intentioned but naïve outsider that you need to have some that make sense to the operational folks and are indicators and you probably set targets and even hold lower level people accountable against (FAA Official Interview A May 16, 2013).

A former FAA planning official describes the interplay between DOT and OMB in crafting the priority goals for the agency and the associated performance indicators:

> The department would then put all of the priority goals together with the other modes and then they would send it over to OMB and OMB would have comments. I don't really remember us changing anything because OMB came back and said they wanted something else. But I do know they (DOT) would go back and forth with OMB (Former FAA Official Interview, May 16, 2013).

A DOT official notes the inherent tension between the agency and the department in the selection of goals and performance indicators:

> There's always tension. An agency will simply think that the department simply wants to check off a box. You know? So they'll say here's something, now go away. I've got better things to do. You know, so we're thinking at a very departmental level, looking kind of long term. Ok, this is a small piece in a long agenda. If this piece drops out, other pieces drop out, this doesn't work. So we try to keep all those things moving at the same time. You know, in a way that helps us achieve overall goals (DOT Official Interview, May 16, 2013).

A significant challenge for planners at DOT is not only selecting the appropriate priority goals for all of the agencies under its control, but also gaining access to appropriate sources of data to provide information on how well the agencies are achieving their goals. Often, there may be agreement on the data to collect, but the agency's systems are not compatible with those at DOT:

> A lot of times, our systems are not … geared to deliver that information. This is partly true. In this day and age there is so much information out there you have to sift through this and say is this meaningful information. Sometimes it's a system restriction, meaning I can't get the information out. Sometimes it's a data quality issue. I can't ensure that the information is going in, I can get the information out, but I can't be positive that the information going in is 100% accurate. Sometimes we can get the same information for different modes and it's different.
>
> We know the data we need to get to and we try to come up with some indicators and then a pathway to get, to ultimately get the good data. So we … get an idea of where some things are kind of going, some indicators, it … gives us some, some analytical information, … a way to infer whether the direction of the specific objective is … going to point us to a specific objective and other instances, sometimes we have to start from scratch (DOT Official Interview, May 16, 2013).

The issue of data quality and a solid baseline for an indicator becomes larger as an agency goal moves up in visibility to external stakeholders, becoming one that is reported as a priority goal metric on OMB's Performance.gov website. In the case of measuring air traffic control safety, there was debate within the agency over what metric would be most appropriate to report publicly on Performance.gov. The FAA's other safety indicators, such as commercial air carrier fatality rate and the general aviation fatality rate, are both salient and relatively straightforward in their computation. Using SRER as a measure of air traffic control safety, particularly given the recent focus on operational errors, would have put the FAA in the position of having to defend a new metric that was not fully developed (FAA Planning Official, May 13, 2013). Therefore, FAA, DOT, and OMB decided to report the percentage of the top five risk categories (see box on p. 26) that were resolved during the course of the year. This indicator allows the FAA to focus on a more risk-based measure of air traffic control errors, while

also focusing on the positive changes made to air traffic as a result of the data collected through its new sources of incident reporting data.

Finally, the air traffic controller union, NATCA, has played a significant role in the transition to the use of incident reporting as an indicator of air traffic safety. Because NATCA encouraged its employees to self-disclose incidents through ATSAP, a valuable source of performance information from frontline employees is now available to the agency to determine its highest-risk losses of separation and to develop corrective actions to mitigate those risks in the future. Additionally, obtaining NATCA buy-in for the increased detection of LoSS through TARP was critical in achieving the full implementation of the system.

While NATCA supported the increased collection of incident reports, the union opposes the merging of ATSAP and CEDAR because of concerns over the possible identification of controllers who commit errors. The inability of the FAA and NATCA to come to agreement on the merging of these two sources of data has led to criticism from both the DOT-IG and GAO and increased scrutiny from Congress and the media.

Lessons Learned

Strategic Lessons

As agencies report more performance information, including incident reporting, there will be increased scrutiny of performance from external stakeholders. The paradox of increased performance information collection and reporting is that as better sources of data and new analytical techniques give a more transparent picture of agency performance, there may be increased scrutiny from outside stakeholders. This can potentially discourage the collection and reporting of such data.

This lesson was quite evident as the FAA Air Traffic Organization (ATO) began to collect voluntary reports from employees through the Air Traffic Safety Action Program (ATSAP) and electronic incident reports through the Traffic Analysis Reporting Program (TARP). As the new sources of information were collected and used to supplement existing performance indicators (based on number of operational errors [OEs]), members of Congress and the media began to question why the number of OEs appeared to drastically increase. Following the greater scrutiny of the OE indicator, the FAA decided to change the air traffic safety indicator to the System Risk Event Rate (SRER), which focuses on limiting the risk of losses of separation rather than the number of OEs.

Agencies must proactively educate key stakeholder groups on new performance indicators and incident reporting systems. A key lesson learned from the FAA ATO case is that it is vital for agency leaders to educate stakeholders, including OMB, members of Congress, GAO, the inspector general, employee unions (if applicable), and key interest groups on the data used to construct performance indicators, along with the data's interpretation.

FAA leaders were often forced to act from a reactive rather than proactive position in explaining the increased number of OEs due to increased detection of incidents through TARP. Because the agency was defending its new risk-based incident reporting systems, it never had a chance to fully explain or promote its new paradigm for measuring air traffic control safety, and the steps it was taking to implement corrective actions to identified safety hazards. Educating, or even involving, stakeholder groups in the development of indicators that use new sources of performance information will help alleviate some of the scrutiny that may result from a perceived drop in performance due to greater data fidelity.

Agencies must develop indicators that are easily understood by external stakeholders. The abandonment of the OE indicator, coupled with the development of the SRER indicator by the FAA ATO, illustrated the need to more closely align the political tenability of a performance metric with the agency's desire to measure its performance through a risk-based metric. SRER allows the FAA to present its new sources of performance data in a meaningful way that evaluates the risk of each serious error detected through TARP. Additionally, the performance information reported by TARP and ATSAP was analyzed by the Risk Analysis Process (RAP) and

ATSAP analysts to develop the top five risk categories within air traffic control. The agency determined that progress on resolving the top five risk categories, rather than the total number of incidents, would be a politically tenable indicator. The use of this indicator as a performance measure for DOT on Performance.gov indicates that the FAA worked closely with DOT and OMB on communicating the importance of the risk-based metric to measuring the success of the ATO safety program.

Management Lessons

In order to be used by agency leaders for decision-making, data analytics processes must focus on reviewing performance information derived from incident reporting systems in a timely manner. One major success of both ATSAP and TARP has been the thoroughness and timeliness of the process for reviewing the data provided through these programs. ATSAP Event Review Committees (ERCs) in each service area meet twice a week to review and validate each narrative ATSAP report. RAP panels in each service area meet three to four times a week to examine the highest-risk incidents reported to the Comprehensive Electronic Data Analysis and Reporting (CEDAR) tool through electronic TARP detection. This frequent validation and assessment of the data provided by these programs allows for continual updates on the ATO's safety performance. Additionally, the validation process allows for constant revisiting of the top five risk categories in air traffic. Finally, the timely validation of risk analysis events through TARP allows for a continual feedback loop on agency performance through the SRER metric.

Collaboration between managers and employees on the use of self-reported incident performance information is critical to success. While the GAO and others have focused on the importance of achieving top-level buy-in for the use of performance information in agencies, the FAA ATO case illustrates the importance of widespread employee buy-in regarding the collection of performance information that can be attributed to them personally. As the FAA developed both its ATSAP and TARP incident reporting systems, obtaining buy-in from the air traffic controllers union (NATCA) was a critical step in allowing the agency to move forward in its shift toward risk-based performance indicators. Because NATCA encouraged its employees to self-disclose incidents through ATSAP, a valuable source of performance information from frontline employees is available to the agency to determine its highest-risk loss of standard separation (LoSS) and develop corrective actions.

Additionally, obtaining NATCA buy-in for increased LoSS detection through TARP was critical in achieving full implementation of the system. The close relationship between NATCA and FAA in advancing these safety improvements also allowed both the union and the agency to present a unified front when answering criticism related to increased operational errors from members of Congress, GAO, the DOT-IG, and the media. However, the GAO and DOT-IG have noted that the inability of the FAA and NATCA to come to an agreement on the merging of ATSAP and TARP data is a major weakness of the FAA's incident reporting system.

Analytical Lessons

Agencies must balance the pressure to develop indicators in a timely manner with the need for rigorous indicators that effectively measure agency performance. In agencies, there is often pressure to move forward without the necessary rigor behind indicators and performance information. This pressure often results from increased public scrutiny, poor performance in achieving agency goals, or as a consequence of congressional authorizations or appropriations. The FAA case illustrates how this pressure can lead agencies to rush to implement underdeveloped indicators that potentially may not be based on reliable and valid information.

The FAA is still in the process of developing baselines and targets for a majority of its measures listed as indicators for success under Destination 2025, due to a lack of resources to collect data on several of the indicators. Additionally, many in the FAA believe that the SRER indicator was rushed as an alternative to the Operational Error Rate indicator in response to increased public scrutiny, without sufficient information on how the data would be analyzed and synthesized. To effectively collect and use performance information for internal operational purposes, agencies must resist the temptation to reactively collect information and implement indicators that are expedient for external stakeholders.

At the same time, agencies must resist the temptation to develop the perfect indicator. Often, specialists within agencies will wait to release information or indicators as they look for more valid and robust sources of data. If agency managers decide to wait for the perfect indicator or source of data, they run the risk of looking indecisive or appearing to obfuscate the release of indicators to external stakeholders. The FAA's experience with SRER illustrates that managers chose to move forward with underdeveloped measures in response to external stakeholders.

Agencies need to develop and invest in rigorous analytical techniques to make the most of new sources of performance information from incident reporting systems. Another lesson from the FAA ATO case is that more data do not necessarily mean more performance information. While the agency has a wealth of valuable data from both ATSAP and TARP, both the GAO and the DOT-IG have identified the lack of analytical tools that merge these sources of data as a weakness in improving air traffic safety. Officials at the FAA noted that the inability to design a risk-assessment tool applicable to both the qualitative and subjective ATSAP reports and the quantitative and unbiased electronic TARP reports was a challenge that the agency was trying to overcome. As the GAO survey of federal managers indicates, the lack of adequate analytical techniques in the FAA is a commonly repeated occurrence across government.

Conclusion

Many executives and managers across a variety of agencies have spent valuable resources to collect and analyze data on how effectively employees are implementing the organization's mission. The outputs and outcome of frontline employees' work can be difficult or impossible to observe without innovative data collection techniques. The focus on data collection on employee performance is particularly important in high-reliability agencies, where poor performance can result in high-impact consequences such as loss of life or large-scale disaster.

While the use and success of incident reporting systems has increased drastically over the past decade, many managers in high-reliability agencies have faced challenges when trying to use reported incidents as indicators of an agency's commitment to performance improvement. The report's lessons learned can be used by senior agency managers to create strategies for developing incident reporting systems, along with strategies for using the data from these systems as performance indicators.

Appendix I: Methodology

This report uses a single descriptive case study approach (Yin 2003) to explore performance management in the Federal Aviation Administration (FAA). Descriptive case studies are useful for explaining phenomena that occur in complex contexts. Given the complexity of the subject of performance management and the varying use of performance information across the federal government, a rich description of one agency's experience with the use of performance data is an appropriate methodology to generate lessons learned and practical recommendations.

The FAA was chosen as the case study for the report for a number of reasons:

- First, the author's experience as a policy analyst at the FAA provided valuable contextual understanding of the environment in which performance management takes place.
- Second, the FAA has a reputation among performance managers as a leader in the area of performance management, particularly in the use of performance information to drive agency action. The FAA was awarded the 2004 Strategic Planning Award from the Association of Strategic Planners for its *Flight Plan* along with the 2004 Certificate of Excellence in Accountability Reporting (CEAR) for its Performance and Accountability report. More recently, both the Center for American Progress (2011) and the Partnership for Public Service (2011) have highlighted the FAA's strategic planning process and the use of data and analytics to improve agency performance as exemplars for other federal agencies.
- Third, the FAA has invested significant agency resources in the collection of operational performance data (U.S. DOT-IG Report AV 2013-046; Mills 2010), making it a good example of how an agency uses data and analytics to improve performance.
- Finally, as an agency of the Department of Transportation (DOT), the FAA is not directly responsible for GPRAMA implementation.

The primary source of data used to develop the FAA case study and the subsequent lessons learned was interviews with officials at FAA's Office of Policy and Plans and ATO (both headquarters and field personnel), current and former FAA strategic planners, DOT Office of the CFO, Office of Management and Budget (OMB), Performance Improvement Council (PIC), and the Government Accountability Office (GAO). In order to receive candid responses to the questions posed, each interviewee was grated anonymity and will be identified only by their organizational affiliation. The primary data was triangulated with other secondary sources of data including FAA Advisory Circulars, FAA Orders, FAA Strategic Plans, FAA Performance and Accountability Reports, Information from the DOT's Performance.gov website, OMB Circular A-11, DOT-Inspector General Reports, and GAO Reports to develop a rich descriptive case study of the FAA's use of data and analytics to improve performance in the ATO.

Appendix II: A Typology of Agencies

Production agencies are those agencies where both outcomes and outputs are easily observable. Wilson (1989) argues that while accountability is often high in production agencies due to the observability of outcomes and outputs by external stakeholders, these agencies are often plagued by a "Gresham's law of information" where measurable outputs and outcomes tend to push out other potential indicators that are not currently measured by the agency. Because of the focus on mission in production agencies, indicators tend to be very closely tied to the central work of the agency (e.g., the number of tax returns filed for the IRS).

A *procedural agency* is one where outputs are observable, but the near-term outcome of the agency's work is harder to measure. Procedural agencies are often marked by close adherence to internal processes and procedures and a means-oriented performance measurement structure: how the employee does their job is more important than whether the work achieves the desired outcome. The means-oriented nature of procedural agencies leads to more metrics focused on employee adherence to the rules through error reporting. Also, many procedural agencies can also be described as having a high risk of a highly visible incident as a result of a breakdown of compliance with internal rules or procedures.

Table A-1: Wilson's Typology of Agencies

	Outcomes Observable	**Outcomes Not Observable**
Outputs Observable	Production Agencies (Examples: Social Security Administration/ Internal Revenue Service)	Procedural Agencies (Example: Occupational Safety and Health Administration/ Federal Aviation Administration's Air Traffic Organization)
Outputs Not Observable	Craft Agencies (Examples: Army Corps of Engineers/Forest Service)	Coping Agencies (Examples: Educators and Police)

Craft agencies are agencies where the outcome of the work is easily observable, but the day-to-day outputs of employees are more difficult to measure. In craft agencies, managers tend to rely on the ethos of service and professionalism of the workforce to ensure that outputs are achieved. Significantly, Wilson (1989) argues that because of this sense of professionalism in craft organizations, managers design performance indicators to measure multiple outcomes.

Coping agencies are agencies where neither outputs nor outcomes are easily achievable. The common examples of coping agencies are local police forces (Whitley 2013) and school teachers. In many coping agencies, there is often a high degree of conflict between managers and frontline employees due to the divergent focus of managers (on constraints from external stakeholders) and those of frontline employees (on constraints from internal stakeholders).

Acknowledgments

The author would like to thank officials at the FAA, OMB, and GAO who agreed to be interviewed for this research project. Your insights and revisions to the report have made it a much stronger piece of work. Additionally, the author would like to thank John Kamensky and Mark Abramson for their excellent suggestions to make the report much stronger. Finally, the author would like to thank his colleagues in the Department of Political Science at Bowling Green State University who read previous drafts of this report.

References

Behn, Robert. 2003. "Why Measure Performance? Different Purposes Require Different Measures" *Public Administration Review.* 63(5): 586–606.

Brass, Clinton. 2012. "Changes to the Government Performance and Results Act: Overview of the New Framework of Products and Processes." Congressional Research Report No. R42379). Washington D.C.: Library of Congress Congressional Research Service.

Cooke, David and Thomas Rohleder. 2006. "Learning from Incidents: From Normal Accidents to High Reliability" *System Dynamics Review.* Vol. 22: 213–239.

DOT Official Interview. May 16, 2013. Washington, D.C.

Ellig, Jerry Maurice McTigue, and Henry Wray (2012) *Government Performance and Results: An Evaluation of GPRA's First Decade*. Boca Raton, FL: CRC Press.

FAA ATO Official Interview A. May 16, 2013. Washington D.C.

FAA ATO Official Interview B. June 4, 2013. Telephone Interview.

FAA Official Interview. May 16, 2013. Washington, D.C.

FAA Planning Official Interview. May 13, 2013. Telephone Interview.

Former FAA Planning Official Interview. May 16, 2013. Washington, D.C.

GAO Interview. May 17, 2013. Washington, D.C.

Halsey III, Ashley. (2013). "FAA: Air Traffic Controllers Responsible for Seven Catastrophic Errors" *Washington Post.* September 12, 2013.

Kamensky, John. "GPRA Modernization Act of 2010 Explained" IBM Center for The Business of Government website: http://www.businessofgovernment.org/blog/business-government/gpra-modernization-act-2010-explained-part-1

Kamensky, John and Albert Morales (2005). *Managing for Results, 2005*. Lanham, MD: Rowman and Littlefield.

Kindy, Kimberly. (2013). "USDA Official Defends Pilot Program for Meat Inspection" *Washington Post.* September 13, 2013.

Kravchuk, Robert and Ronald Schack. 1996. "Designing Effective Performance-Measurement Systems under the Government Performance and Results Act of 1993" *Public Administration Review.* 56(4): 348–358.

Levin, Alan. (2013). "Close Calls in U.S. More Than Doubled Last Year, FAA Says" *Bloomberg News.* September 12, 2013.

Long, Edward and Aimee Franklin. 2004. "The Paradox of Implementing the Government Performance and Results Act: Top-Down Direction for Bottom-Up Implementation" *Public Administration Review.* 309–319.

Mills, Russell W. 2010. *Voluntary Regulatory Partnership Programs with Industry: Lessons from the Federal Aviation Administration.* Washington D.C.: IBM Center for The Business of Government.

Moynihan, Donald P. 2008. *The Dynamics of Performance Management: Constructing Information and Reform*. Washington D.C.: Georgetown University Press.

Moynihan, Donald P. 2013. *The New Federal Performance System: Implementing the GPRA Modernization Act.* Washington D.C.: IBM Center for The Business of Government.

Moynihan, Donald P. and Stéphane Lavertu. 2012. "Does Involvement in Performance Reforms Encourage Performance Information Use? Evaluating GPRA and PART" *Public Administration Review* 72(4): 592–602.

Nicholson Crotty, Sean, Nick Theobald, and Jill Nicholson Crotty. 2006. "Disparate Measures: Public Managers and Performance-Measurement Strategies" *Public Administration Review.* 66(1): 101–113.

OMB/PIC Interview. May 17, 2013. Washington, D.C.

U.S. Department of Transportation, Office of the Inspector General. 2013. FAA's Efforts to Track and Mitigate Air Traffic Losses of Separation are Limited By Data Collection and Implementation Challenges. AV-2013-046. Washington, D.C. DOT.

U.S. Department of Transportation, Office of the Inspector General. 2012. Long Term Success of ATSAP Will Require Improvements in Oversight, Accountability, and Transparency. AV-2012-152. Washington, D.C. DOT.

U.S. Department of Transportation, Office of the Inspector General. 2004. Controls Over the Reporting of Operational Errors AV-2004-085. Washington, D.C. DOT.

U.S. Department of Transportation, Office of the Inspector General. 2000. Actions to Reduce Operational Errors and Deviations Have Not Been Effective. AV-2001-11. Washington, D.C. DOT.

U.S. Federal Aviation Administration. 2013. Air Traffic Organization. Excellence Through Knowledge. FY 2012 Safety Briefing.

U.S. Federal Aviation Administration. 2012. Air Traffic Organization. A Plan for the Future 2012–2021.

U.S. Federal Aviation Administration. 2012. Air Traffic Organization Occurrence Reporting. FAA Order JO 7210.632.

U.S. Federal Aviation Administration. 2012. Air Traffic Organization Quality Assurance Program. FAA Order JO 7210.633.

U.S. Federal Aviation Administration. 2002. Air Traffic Quality Assurance. FAA Order JO 7210.56C.

U.S. Federal Aviation Administration. 2012. Air Traffic Organization Quality Control. FAA Order JO 7210.634

U.S. Federal Aviation Administration. 2011. *Destination 2025.*

U.S. Federal Aviation Administration. 2013. FY 2012 Performance and Accountability Report.

U.S. Federal Aviation Administration. 2012. FY 2011 Performance and Accountability Report.

U.S. Federal Aviation Administration. 2010. FY 2009 Performance and Accountability Report.

U.S. Federal Aviation Administration. 2009. FY 2008 Performance and Accountability Report.

U.S. Federal Aviation Administration. 2003. FY 2002 Performance and Accountability Report.

U.S. Federal Aviation Administration. 2002. FY 2001 Performance and Accountability Report.

U.S. Federal Aviation Administration. 2001. FY 2000 Performance and Accountability Report.

U.S. Federal Aviation Administration. 2000. FY 1999 Performance and Accountability Report.

U.S. Federal Aviation Administration. 2012. Voluntary Safety Reporting Programs. FAA Order JO 7200.20.

U.S. Government Accountability Office. 2013. Executive Branch Should More Fully Implement the GPRA Modernization Act to Address Pressing Governance Challenges. GAO-13-518. Washington, D.C.: GAO.

U.S. Government Accountability Office. 2013. Data-Driven Performance Reviews Show Promise But Agencies Should Explore How to Involve Other Relevant Agencies. GAO-13-228. Washington, D.C.: GAO.

U.S. Government Accountability Office. 2012. FAA Is Taking Steps to Improve Data, but Challenges for Managing Safety Risks Remain. GAO-12-660T. Washington, D.C.: GAO.

U.S. Government Accountability Office. 2011. Enhanced Oversight and Improved Availability of Risk-Based Data Could Further Improve Safety. GAO-12-24. Washington, D.C.: GAO.

U.S. Government Accountability Office. 2008. Government Performance: Lessons Learned for the Next Administration on Using Performance Information to Improve Results. GAO-08-1026T. Washington, D.C.: GAO.

U.S. Government Accountability Office. 2004. GPRA Has Established a Solid Foundation for Achieving Greater Results. GAO-04-38. Washington, D.C.: GAO.

U.S. Office of Management and Budget (OMB). 2013. The President's Budget for Fiscal Year 2014: Analytical Perspectives. Washington D.C.: Government Printing Office.

U.S. Office of Management and Budget (OMB). 2013. OMB Circular A-11: Preparation, Submission, and Execution of the Budget. Washington D.C.: Government Printing Office.

U.S. Senate Committee on Homeland Security and Governmental Affairs. 2010. GPRA Modernization Act of 2010. Washington D.C.: U.S. Government Printing Office.

Wilson, James Q. 1989. *Bureaucracy: What Government Agencies Do and Why They Do It.* Basic Books.

Yin, Robert K. 2003. *Case Study Research: Design and Method*. Newbury Park: Sage Publications.

About the Author

Russell W. Mills, Ph.D. is an Assistant Professor in the Department of Political Science at Bowling Green State University. Prior to joining the faculty at BGSU, Mills worked as a policy analyst in the Office of Aviation Policy and Plans at the Federal Aviation Administration in Washington, D.C., from 2011 to 2012. During his tenure at the FAA, Mills assisted in the development and analysis of reauthorization and appropriations legislation, conducted analysis of the Federal Contract Tower Program, developed and analyzed policy proposals related to NextGen implementation, and analyzed the implications of tax proposals to the Airport and Airway Trust Fund.

Mills' research interests include federal aviation policy, local economic development efforts, voluntary self-disclosure programs, performance management, and governance networks. His 2011 co-authored article titled "Accountability in Governance Networks: Lessons From Hurricane Katrina" won the 2012 Marshall Dimock Award for the best lead article in *Public Administration Review.* Additionally, Mills published an IBM Center for The Business of Government report in 2010 titled *Voluntary Regulatory Partnership Programs with Industry: Lessons from the Federal Aviation Administration.*

Mills holds a PhD in Political Science from Kent State University, a Master of Public Administration degree from the University of Vermont, and a BA in Political Science from Westminster College in New Wilmington, PA.

Key Contact Information

To contact the author:

Russell W. Mills, Ph.D.
Assistant Professor
Department of Political Science
Bowling Green State University
124 Williams Hall
Bowling Green, OH 44012
(419) 372-7329

e-mail: millsrw@bgsu.edu

Reports from IBM Center for The Business of Government

For a full listing of IBM Center publications, visit the Center's website at **www.businessofgovernment.org.**

Recent reports available on the website include:

Acquisition

A Guide for Agency Leaders on Federal Acquisition: Major Challenges Facing Government by Trevor L. Brown
Controlling Federal Spending by Managing the Long Tail of Procurement by David C. Wyld

Assessing the Recovery Act

Recovery Act Transparency: Learning from States' Experience by Francisca M. Rojas
Key Actions That Contribute to Successful Program Implementation: Lessons from the Recovery Act by Richard Callahan, Sandra O. Archibald, Kay A. Sterner, and H. Brinton Milward
Managing Recovery: An Insider's View by G. Edward DeSeve
Virginia's Implementation of the American Recovery and Reinvestment Act: Forging a New Intergovernmental Partnership by Anne Khademian and Sang Choi

Collaborating Across Boundaries

Collaboration Between Government and Outreach Organizations: A Case Study of the Department of Veterans Affairs by Lael R. Keiser and Susan M. Miller
Using Crowdsourcing In Government by Daren C. Brabham
Developing Senior Executive Capabilities to Address National Priorities by Bruce T. Barkley, Sr.
Beyond Citizen Engagement: Involving the Public in Co-Delivering Government Services by P. K. Kannan and Ai-Mei Chang
Implementing Cross-Agency Collaboration: A Guide for Federal Managers by Jane Fountain

Fostering Transparency and Democracy

Assessing Public Participation in an Open Government Era: A Review of Federal Agency Plans by Carolyn J. Lukensmeyer, Joe Goldman, and David Stern

Improving Performance

Predictive Policing: Preventing Crime with Data and Analytics by Jennifer Bachner
The New Federal Performance System: Implementing the GPRA Modernization Act by Donald Moynihan
The Costs of Budget Uncertainty: Analyzing the Impact of Late Appropriations by Philip G. Joyce

Using Technology

Rulemaking 2.0: Understanding and Getting Better Public Participation by Cynthia R. Farina and Mary J. Newhart
The Use of Data Visualization in Government by Genie Stowers
Mitigating Risks in the Application of Cloud Computing in Law Enforcement by Paul Wormeli
Challenge.gov: Using Competitions and Awards to Spur Innovation by Kevin C. Desouza
Working the Network: A Manager's Guide for Using Twitter in Government by Ines Mergel

IBM Center for
The Business of Government

About the IBM Center for The Business of Government

Through research stipends and events, the IBM Center for The Business of Government stimulates research and facilitates discussion of new approaches to improving the effectiveness of government at the federal, state, local, and international levels.

About IBM Global Business Services

With consultants and professional staff in more than 160 countries globally, IBM Global Business Services is the world's largest consulting services organization. IBM Global Business Services provides clients with business process and industry expertise, a deep understanding of technology solutions that address specific industry issues, and the ability to design, build, and run those solutions in a way that delivers bottom-line value. To learn more visit: ibm.com

For more information:
Daniel J. Chenok
Executive Director
IBM Center for The Business of Government
600 14th Street NW
Second Floor
Washington, DC 20005
202-551-9342
website: www.businessofgovernment.org
e-mail: businessofgovernment@us.ibm.com

Stay connected with the IBM Center on:

or, send us your name and e-mail to receive our newsletters.