STANDARDIZING PERFORMANCE METRICS

5th USA /Europe Air Traffic Management R&D Seminar, June 2003

Tamara Breunig, CSSI Inc., Washington DC, USA

Steve Bradford, Federal Aviation Administration, Washington DC, USA

Diana Liang, Federal Aviation Administration, Washington DC, USA

Abstract

"Businesses that succeed and make money constantly assess themselves and improve in all dimensions of their business; metrics are the cornerstone of their assessment, and the foundation …" [1]

Over the past decade, performance metrics have been a cornerstone of private industry performance. Industry has focused strongly on measurement -- how to measure the performance of operations, systems, employees, and strategic goals.

As the importance of measurement increased, various methodologies were created to offer a foundation for assessment and a means for comparison within each industry. Methodologies such as "balanced scorecard" or "six sigma" became the bible of performance measurement. Each method offered a detailed framework on the measurement process by clearly defining the objectives, metrics (or indicators), and means of calculating. Certification in a particular method could be achieved through formal training and implementation of the framework standards. Certification benefited all involved; it enabled a business to be grouped into a separate category, generally considered to uphold higher expectations and level of quality. It also allowed for the business to be compared on a consistent basis with all other certified businesses, even globally.

The purpose of this paper is to show the need for implementing a standard methodology for the governing aviation community – a "balanced scorecard" – for performing assessment of systems and concepts.

Identifying the Need

Performance measurement is considered critical in making and supporting business decisions. It provides a clear and objective way of measuring success. However, performance measurement is an essential yet confusing aspect of many industries. It carries a unique definition dependent on the industry/community.

In private industry, performance measurement relates to the level of profit obtained while providing a service or product. The level of profit obtained is an indicator of success and is measured by variables such as "return on investment (ROI) and productivity, or revenues minus total cost of ownership (TCO)" [2]. These figures are carefully defined and are all measured the same across the industry, so it is easy for an individual to determine each business's net worth and quickly evaluate its performance at any time.

Likewise, in academia, performance measurement primarily relates to the level of education obtained while attending school. The level of education obtained is an indicator of success and is measured by standard grading systems and tests, such as the Standard Aptitude Test (SAT). These types of tests are designed and regularly evaluated by a board of experts to ensure the quality and accuracy of the test. Through implementing standardized tests, universities are able to identify a student's level of performance with ease. Although standardized tests are only one aspect of performance measurement in academia, other types of measurement methods include grade point average (GPA), extra curricular activity, community involvement, etc.

Governing organizations also have definitions of performance measurement. However, their existence is not always measured in a financial nature, but rather on mission effectiveness and the opinions of the local/global community. In the United States for example, the representatives in Congress passed the Government Performance and Results Act (GPRA) back in 1993, requiring federal agencies to establish a plan with goals and methods of measuring against those goals. Since the act was formed, the world of performance measurement has expanded, adding new metrics that depend on the focus of the agency.

Similar to the examples above, performance measurement for governing aviation agencies has relied on a formal, yet informal framework of measurement. The initial framework consisted of the historic objectives, Safety and Delay; both were established to address the concerns of consumers and the general public. Under these objectives there are abundant metrics associated with each.

Reviewing the objective of Safety, its defining metrics have spanned from the original metric of Number of Accidents to new metrics such as Number of Conflicts. This expansion is caused by further defining objectives and continuing research aimed at determining the cause and effect.

For example, if the goal is to reduce the accident rate, factors affecting the safety of an aircraft must be measured. Rather than simply relying on the original metric of Number of Accidents to evaluate the situation, new, more-defined metrics have been developed that relate more specifically to potential accident factors, such as the aforementioned metric Number of Conflicts.

However, with the expansion of aviation metrics and the lack of standard methodology that is present in industry and academia, each metric has the potential to be measured with different interpretations and calculations. A quick review of the metric Number of Conflicts could be interpreted as the number of *actual* conflicts, conflicts that were presented to the controller for resolution. Whereas another reader might interpret the metric to mean the number of conflicts that would occur given the original flight plan was flown with no adjustments, hence the number of *potential* conflicts. This illustration of misinterpreting the metrics presented occurs frequently due to the complexity of measuring activities in the airspace and the lack of standardized definitions for the metrics. By standardizing these definitions, this problem could be easily avoided.

In addition to misinterpreting the meaning behind a metric, duplicate metrics are appearing as performance measurement grows. Duplicate metrics are sometimes difficult to uncover without reading into the definition or purpose of the metric. Essentially, two metrics may have the same or nearly the same definition but different titles or names.

These are just a few cases justifying the need for a formal framework or balanced scorecard. As the demand for aviation services increases, the system and operational challenges will increase along with the need for metrics to measure the new developments in the aviation industry.

To further understand and validate the need for a standard performance measurement framework in the aviation community, research was conducted in the area of performance metrics, key performance indicators, human or system performance metrics, etc. The remainder of the paper provides details explaining the current efforts towards standardization and the discrepancies encountered from various studies, experiments, and white papers discussing performance metrics.

Efforts Towards Standardization

After each analysis or performance review, familiar statements such as “20 percent increase in capacity”, or “5 percent fuel savings”, or “15 percent delay reduction” are often used to describe the final results for the assessment of the new system or concept. The final results inform the audience of the high-level benefits, yet leave many unanswered questions. How were the benefits calculated? What sorts of metrics were employed to generate the benefits? How accurate are these benefits? How does this new system compare to the current system? How does this new system compare to our global partners’ systems? All these questions and more could be answered by simply providing a summary of the certified framework applied during the study, with references to any changes in the methodology or metrics.

Instead, many of the questions can only be answered by reading the rather lengthy report behind the final answers, which details the methodology, metrics, and calculations applied to that particular study. However, even after reading and understanding the implied methodologies of one study, it is typically a challenging task to attempt a comparison of results with other studies. Researchers and analysts are finding that “many [of the] approaches used in past…has meant it has been impossible to compare their results and conclusions, and thereby identify the best future operational concept …” [3]. This problem of varying methodologies and approaches is a growing concern in the aviation community.

The Framework

In an attempt to resolve the varying methods of performance measurement, a number of approaches to standardizing performance measurements have been constructed, resulting in an assortment of frameworks (such as, ISO 13236). Essentially each framework presented consists of a hierarchy format shown in the figure below, Figure 1.

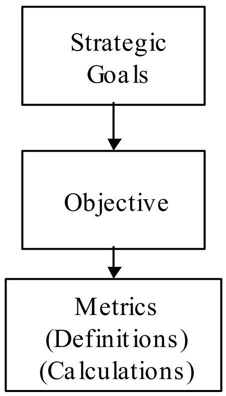

Figure 1. Basic Framework

The basic framework begins with the top layer, where the strategic goals or the scope of the project is defined. This layer is then mapped to the corresponding objective or objectives. The objectives are generally used to illustrate to the stakeholders the level of success achieved in reaching the outlined goals.

In order to measure the level of success, the objectives are linked to the bottom layer, the metrics. Included in the metric layer is the definition and any necessary calculation guidelines or formulas.

Although each framework approach is broadly similar, the end results are not consistent across studies and are considerably different in taxonomy [3]. The discontinuity between approaches begins right in the heart of the framework, the objective, which is the basis for identifying measurable factors.

Objective Layer

The word "objective" is commonly used for this layer, although several other studies refer to this layer as Generic Characteristics or Metric Categories. However, the difference in naming conventions for this layer is of minor concern; the real problem lies in determining the list of objectives.

Table 1, found on the following page, shows a list of objectives identified. This list is an extract from various reports, generally authored by Eurocontrol, the National Aeronautical Space Administration (NASA), and the Federal Aviation Administration (FAA). Each report was produced to detail the proposed structure of performance measurement within the agency and/or validate previous experiments performed and the measurement applied during the experiment. The table shows the variety of terms used to define the objective layer.

The objective terms are listed in alphabetical order and the references are listed in chronological order corresponding to the number found in the Reference section. In each cell, an "X" indicates that the term was listed as an objective in the framework referenced.

It should be noted that some terms were actually duplicates, even though the terminology differed. One example is the objective Role of Human in ATM, which is equivalent to Human Involvement. Although a clear definition was not provided for the two objectives, they were determined to be equivalents based off of the metrics associated with each objective. The Role of Human in ATM is an objective established by the stakeholders to set forth that "humans remain an essential part of the aviation system" [12]. One example of a metric in this area is evaluating the "effect of shifting separation authority on controller and pilot workload and situation awareness activities" [12]. In summary the metrics relate to activities where humans are involved, thus Human Involvement.

Another area to note is the number of terms associated with measuring human performance. Human performance metrics are considered "a less direct measure" [3], and therefore are not listed in the table of objectives. For example, the human performance concept of workload is "closely linked to [the objective of] capacity because workload is generally recognized as a limiting factor for system capacity" [3]. These terms identified as lower layers are as follows:

- Workload
- Situation Awareness
- System Monitoring
- Teamwork
- Trust
- User Acceptance
- Human Error

Aside from the concern of what constitutes as an objective, the framework structure is in question too. The last column of the table is labeled with a "PA", which stands for "performance area", a term described as the middle link between the objective and a metric. One example of a performance area is “flight efficiency”, which has an impact on and therefore links to the objective “efficiency” and the objective “environment”.

The “X” under the PA column indicates that the term referenced has been categorized as a performance area; however, note that some performance areas have been considered objectives. For example, Access is listed as a performance area and could be linked to the objective Capacity or Flexibility.

Table 1. Consolidated List of Objectives/Performance Areas

Objectives	3	4	5	6	7-1	7-2	7-3	9	10	12	13	14	PA
Access			X		X	X		X	X				X
ATM Community Participation					X								
Availability						X							X
Capacity	X	X	X	X	X				X	X	X	X	
Civil Military Cooperation													X
Cost Effectiveness					X	X	X						X
Delay						X		X	X				X
Economics										X		X	
Efficiency	X	X	X	X	X			X	X		X		
Environment	X	X	X		X	X	X		X	X	X		X
Equity					X	X							X
Expectations					X								
Flexibility			X	X	X	X		X	X				X
Flight Efficiency						X							X
Global Interoperability					X		X						
Human Involvement										X			
Predictability			X	X	X	X		X	X				X
Productivity									X				
Military Access													X
Quality										X			
Safety	X	X	X	X	X	X	X	X	X	X	X	X	X
Security and Defense	X	X			X		X			X			
Security of Air Transport													X
System Operational Performance							X						
Transition							X						
Uniformity										X			
Expectations													

Reference 7. 7-1. Expectations of the ATM Community, 7-2. EUROCONTROL PRC’s European ATM Performance Measurement System, 7-3. Boeing WTT Report-Phase I – System Performance Requirements

Another example of a performance area in Table 1 is the term Military Access. This could be considered an objective, or could be the performance area linked to the objectives Access and Security and Defense. Of course this assumes that both of these terms are considered objectives. While in the Cooperative Actions of Research and developments in Eurocontrol (CARE) framework, Access is not considered an objective, but a performance area. This type of inconsistency in framework layers and terminology leads to confusion rather quickly. Figure 2 below is an updated version of Figure 1 that better illustrates the relationship between objectives, performance areas, and metrics.

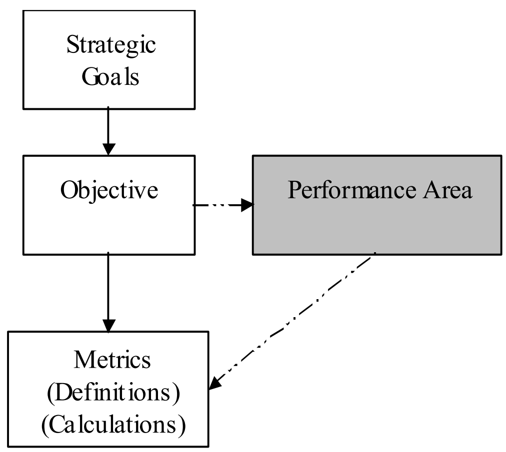

Figure 2. Expanded Framework

Thus far the discussion on the objective layer has shown that the framework can be found with a link directly to the metrics layer or through the sub layer of performance areas.

Another twist in the framework is a method that substitutes the layer of performance area for a "quadrant" layer. This type of framework is exhibited in the FAA's Operational Evolution Plan (OEP), which aims to describe the performance metrics and measurement methodologies applied by the FAA. The framework identifies the FAA strategic goal of meeting the air transportation demands through increasing capacity and decreasing delays, while ensuring safe and secure environment for all [6]. From this, the objectives are defined and placed in a "quadrant" format.

The quadrants represent the four core areas where improvement is needed. These areas are Arrival/Departure, En Route, Airport Weather, and En Route Weather. Underneath each quadrant is a list of metrics that help define the quadrant, thus the metrics layer.

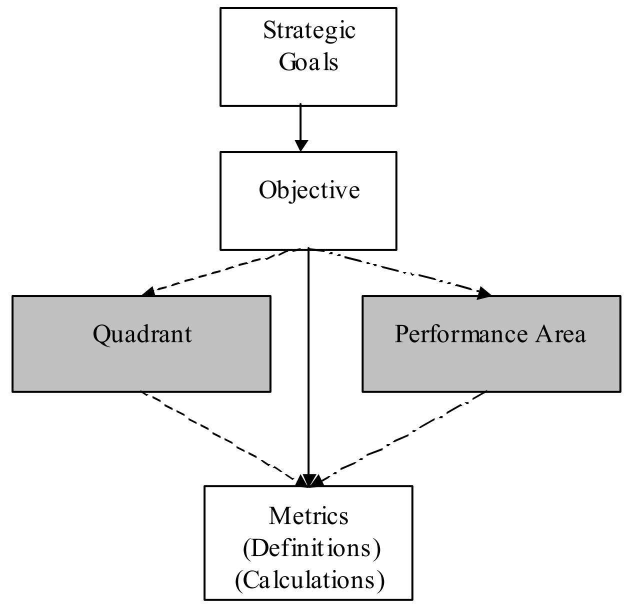

Figure 3. Expanded Framework

Metric Layer

The confusion and differences found in the objective layer continue to be present during the analysis of the metric layer of the framework. The word "metric" can be defined as "the means of measuring benefits against objectives" [3]. Or as a "means of measuring, quantitatively or qualitatively, the different parameters linked to a system, a service, etc." [14]. Essentially all approaches conclude that a metric is a type of measurement used to illustrate an objective or performance area. However, a number of approaches move past the metric layer and identify a few additional layers of categorization.

These additional sub-layers generally relate to classifying a metric by the audience and/or the system phase. The audience describes whom the interested parties or stakeholders would be in relation to the metric. Such parties or stakeholders would include pilots, air traffic controllers, passengers, airlines, cargo, governing agencies, etc. The second layer of system varies according to framework. A multitude of phrases are used to describe this area; a few samples are:

- internal and external [12]

For the layer of metrics, internal and external are the proposed terms found in a Eurocontrol report. The internal metrics provide information on system adjustments or needed refinements of a process, whereas the external metrics provide a means of measuring the overall concept or phases of the system.

- domain or tool specific [5]

According to a framework presented by the NASA, the metrics layer is spilt into domain-specific metrics followed by tool-specific metrics. Domain metrics are those that "address impacts of a group of tools [or system] within each of…the domains" [5]. There are four domains: surface, terminal, en route, and airborne. Following this layer are the tool-specific metrics. This layer consists of metrics that are applicable for a particular tool.

- macro, meso, and micro [6]

Similar to the NASA framework, the framework established by the FAA's OEP consists of layers referring to the domains and tools. The framework has a three-level hierarchy: macro, meso, and micro. Starting from the bottom, a micro metric is one that describes the effects on a location or specific component of a program/system. A micro metric would be utilized in measuring the impact of a new runway at a specific airport such as Ronald Reagan Washington National. The middle layer, meso metrics, is used to describe effects on specific regions, systems groups, or areas of interest. A meso metric would be used to measure the impact of reducing the vertical separation minimum in the Northeast corridor. Finally, the top level, macro metrics, covers the NAS-wide metrics. These metrics are used to measure the overall NAS health and performance, such as percent of flights on time.

In addition to the metric sub layers already discussed, a unique sub-layer was briefly mentioned in the CARE approach. While establishing the framework for this particular approach, the layer "measure" was encountered during research. Measure defines the specific data item that quantifies a metric. The layer was not added to the final framework due to the uncertainty and similarity, "measure forms part of the definition of a metric and therefore does not constitute a further layer in the hierarchy" [3].

The final component of the framework is the definition and calculation of the metrics, a process just as important as the other layers/sub-layers of the framework. As discussed under the section *Identifying the Need* with the metric Number of Conflicts, the lack of a clear definition and calculation is potential for misinterpreted results.

Problems often arise when consensus is not reached around the specific definition of each metric; such is the case when working with the metric Daily Traffic Count. On the surface this metric appears to be self explanatory, the total number of flights that occur in a given calendar day. Is it assumed that the count only includes IFR flights? How one defines the number of flights is not commonly shared across organizations. For example, one definition of the metric is the total number of IFR flights that enter the airspace. In this definition, flights considered over flights are counted every time they enter the airspace, even if they are re-entering the airspace. Over flights are those flights that do not depart or arrive at an airport within the controller airspace, but do fly through the controlled airspace.

The last matter to address is the importance of establishing proper links between objectives and metrics. By missing one or two links the study results could end up with no results or even skewed results. For instance, when determining the effects of a new system or concept on sector capacity a common metric is throughput. Throughput is generally calculated as the number of aircraft to traverse a sector in a given window of time. However, looking only at throughput will not tell the entire story. Additional metrics to link include system health, weather, delays, and traffic demand feeding the sector. These metrics help clarify the changes in sector capacity, but are still missing the ability to perform a comparison across all sectors. To be able to compare one sector with another, metrics relating to the sector characteristic need to be linked. These metrics describe the complexity of the sector – the density, number of potential conflicts, resolution of the conflict (time and maneuvers required), etc. Incorporating all of these metrics will relay a truer story about sector capacity and the impact of a new system/concept.

Summary of the Framework

Combining all the methodologies discussed, an updated framework is presented in Figure 4 to further illustrate the differences. The framework has been expanded to include all the layer options discussed. It also includes another sub-layer, subjective or objective. This categorization was found in the majority of the approaches and defined in a consistent manner, where a subjective metric is a judgmental measurement (such as surveys) and an objective metric is a quantitative assessment (such as minutes of delay).

Although each concept and structure is similar, there is enough irregularity in terminology and sub-layers to still require further work on establishing a standard framework.

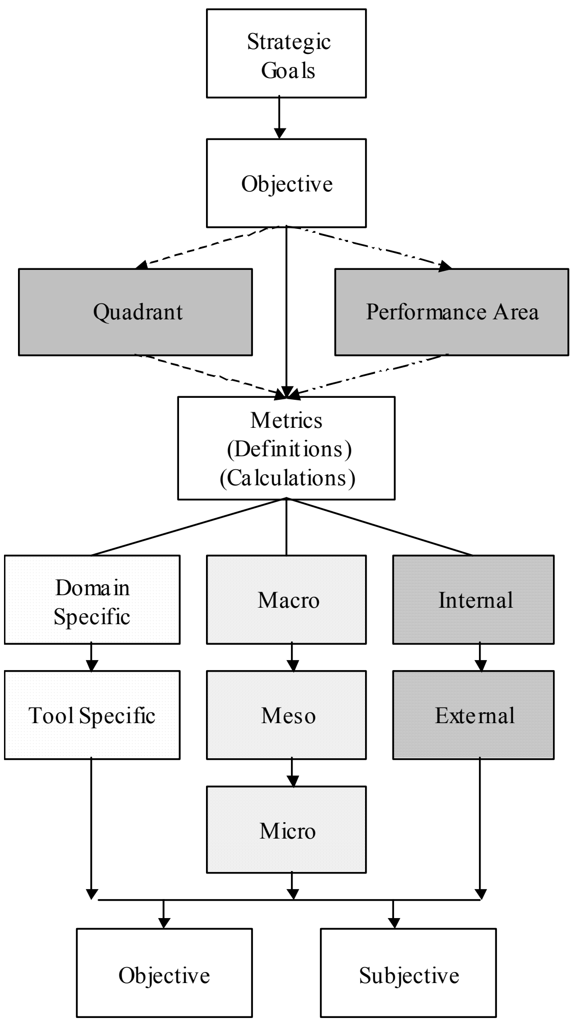

Figure 4. Summarized Framework

Performance metrics play a key role in producing optimal solutions for current and future airspace challenges. They are an integral part in providing feedback on how current operations compare to strategic goals, and provide insight to understanding where various procedural and technological developments are needed.

Global Effort

The complexity of measuring the benefits of a new system or concept is far greater then simply selecting a handful of metrics. However, having the option to refer to past experiences and a certified framework of metrics can reduce the complexity.

It goes without saying that performance measurements are an integral part of the aviation industry. In order for the industry as a whole continue to grow and be successful in future generations, standards must be established. But who should be responsible for establishing the standards?

The ideal solution would be to have a global framework, a methodology created and maintained by a consortium of leaders from the aviation community worldwide. The committee would be responsible for collecting and consolidating all forms of measurements from the industry, including airline carriers, cargo, airports, governing agencies, etc.

The ideal solution would include the creation of an online repository for metrics, a place where analysts, researchers, and management could submit metrics for review by the committee. The repository would establish a central location for inquiring and understanding the current metrics at hand.

By implementing such a system, analysts would be able to quickly reference the methodology behind the "10 minute reduction in delay." The detailed steps on how to calculate the metrics would be in place and certified. In short, the repository would be a key item to use during an assessment due to the credibility associated with it.

Conclusion

Until the community recognizes the need for standardization of performance measurement, the short-term solution is to start moving toward the process.

The FAA is working with the aviation community to establish a formal framework – defining objectives and their correlating metrics, along with the means of calculating the metric. This framework is a work in progress and after completion will still periodically require adjustments as the airspace environment changes. However, it provides a balanced scorecard for reporting the current environment and system performance, and aides in projecting the benefits of new systems and operational concepts.

Key Words

Metrics, Performance Measurement (or Metrics), Measurement (or Metrics) Framework, Key Performance Indicators

References

[1] Trimble, Dave, *How to Measure Success: Uncovering The Secrets Of Effective Metrics* http://www.prosci.com/metrics.htm

[2] Averson, Paul, 1998, *Measuring Performance in Government Research Agencies*, The Balanced Scorecard Institute, http://www.balancedscorecard.org/metrics/govt_labs.htm

[3] Sharples, Michael, John Bennett, 2002, *CARE ASAS Ativity 2: WP2, System Performance Metrics Report, Version 1.0*

[4] 2002, *CARE ASAS Ativity 2: Validation Framework WP3 Deliverable 5, Human Performance Metrics, Version 0.1*

[5] Synder, Phillip, Maria Vellara Skvaicus, James Poage 1999, *Technical Performance Metrics Description Document for the NASA Advanced Air Transportation Technologies Project,* Draft Report

[6] Federal Aviation Administration, 2002, *OEP Metrics Plan,* Pre-Coordination Draft, Version 1.0

[7] EURONCONTROL/Sofreavia, Bui, N., M. Delarche, E. Grasser, B. Miaillier, 2002, *Interim Report on RTSP Modeling*, Version 11, http://www.sofreavia.fr

[8] Alliot, Jean-Marc, Jean-Marc Pomeret, 2002, *Arithmetic Simulation and Performance Metrics*, IEEE

[9] Bradford, Steve, Dave Knorr, Diana Liang, 2000, *Performance Measures for Future Architecture, 3rd* USA/EUROPE ATM R&D Seminar

[10] Federal Aviation Administration, *Airspace Management Handbook- Metrics,* Department of Transportation, Washington, DC.

[11] Bradford, Steve, Lee Brown, Michelle Blucher, 1998, *Use of Performance Metrics in Airspace Systems: US Perspective,* 2nd USA/EUROPE ATM R&D Seminar

[12] EUROCONTROL, SCS Unit – Project CEAMAC, Arino, Thierry, Beatrice Raynaud, Nathalie De Beler, Christian Aveneau, 2002, *Comparison of EACAC and AGIE metrics using CARE/ASAS #2 Framework, Action Plan 5*

[13] Saez, Francisco, Fernando Gomez, John Polky, 2002, *Are Homeogeneous Performance Metrics for the whole ATM system life cycle available?, FAA/EUROCONTROL AP2, 5 and 6 TIM9*

[14] TORCH, 2000, Indicators and Metrics for the Assessment, Memo

Biography

Tamara Breunig is an analyst for CSSI Inc, where her skills are applied to the Advanced Programs group and Oceanic Program Management Business unit. Her responsibilities include researching performance metrics, developing technical capabilities to produce metrics, and performing benefit assessments for future concepts/systems. Prior to CSSI, Ms. Breunig served as a manpower performance analyst for US Airways, evaluating performance of airport stations and collecting workload/passenger profile metrics. In addition, she participated as a member of the CF6 General Electric Aircraft Engine team, providing program and technical support. Ms. Breunig holds B.S. in Industrial Systems Engineering and currently working towards her M.S. in Civil Engineering, Advanced Transportation Systems.

Steve Bradford is the chief scientist for Architecture and NAS Development in the FAA's Office of System Architecture and Investment Analysis (ASD). In this role he has participated in the development of the RTCA NAS Operational Concept and the ICAO ATMCP Global Concept. His organization is also responsible for leading the effort to validate the future concepts, develop the FAA's ATC Information Architecture and leads several co-operative efforts via action plans with Eurocontrol. Prior to his current position, Mr. Bradford was the Manager of the NAS Concept Development Branch. Earlier, Mr. Bradford was lead on several simulation and analytic software development efforts, and conducted early analysis of Free Flight Concepts. From 1987 to 1991 he worked for CACI, Inc. where he led the SIMMOD model development and taught simulation language and modeling courses. He has also worked for the US Navy developing logistic planning models.

Diana Liang works for the Office of System Architecture and Investment Analysis for the Architecture and System Engineering Division. She is responsible for the development of the NAS Architecture Tool and Interface called CATS-I, directing analyses in support of NAS Concept Validation, and the development of Modeling Tools and Fast-Time Simulations to support that validation. This work includes several models she is developing jointly with NASA and cooperative efforts with Europe via Eurocontrol. Prior to working for ASD, Ms. Liang worked in the Office of Energy and Environment for two years as the lead for the Emissions and Dispersion Modeling System (EDMS), updated the FAA's Air Quality Handbook and reviewed Environmental Impact Statements related to emissions. Ms. Liang holds a BS in Computer.