The Evolution of System Safety at NASA

Homayoon Dezfuli, Ph.D; NASA Headquarters; Washington DC, USA
Chris Everett; Information Systems Laboratories; New York, USA
Frank Groen, Ph.D; NASA Headquarters; Washington DC, USA

Keywords: NASA, system safety, safety case, safety assurance

Abstract
The NASA system safety framework is in the process of change, motivated by the desire to promote an objectives-driven approach to system safety that explicitly focuses system safety efforts on system-level safety performance, and serves to unify, in a purposeful manner, safety-related activities that otherwise might be done in a way that results in gaps, redundancies, or unnecessary work. An objectives-driven approach to system safety affords more flexibility to determine, on a system-specific basis, the means by which adequate safety is achieved and verified. Such flexibility and efficiency is becoming increasingly important in the face of evolving engineering modalities and acquisition models, where, for example, NASA will increasingly rely on commercial providers for transportation services to low-earth-orbit. A key element of this objectives-driven approach is the use of the risk-informed safety case (RISC): a structured argument, supported by a body of evidence, that provides a compelling, comprehensible and valid case that a system is or will be adequately safe for a given application in a given environment. The RISC addresses each of the objectives defined for the system, providing a rational basis for making informed risk acceptance decisions at relevant decision points in the system life cycle.

Introduction
The NASA system safety framework is in the process of change. A major motivation for this change is the desire to promote an objectives-driven approach to system safety that explicitly focuses system safety efforts on system-level safety performance, and serves to unify, in a purposeful manner, safety-related activities that otherwise might be done in a way that results in gaps, redundancies, or unnecessary work. An objectives-driven approach to system safety affords more flexibility to determine, on a system-specific basis, the means by which adequate safety is achieved and verified. Such flexibility and efficiency is becoming increasingly important in the face of evolving engineering modalities and acquisition models, where, for example, NASA will increasingly rely on commercial providers for transportation services to low-earth-orbit. A key element of this objectives-driven approach is the use of the risk-informed safety case (RISC): a structured argument, supported by a body of evidence, that provides a compelling, comprehensible and valid case that a system is or will be adequately safe for a given application in a given environment. The RISC addresses each of the objectives defined for the system, providing a rational basis for making informed risk acceptance decisions at relevant decision points in the system life cycle. The RISC is not an add-on to today's system safety practices; it is a means of organizing existing, often disparate, system safety products such as failure modes and effects analyses (FMEAs), hazard analyses (HAs), and probabilistic risk assessments (PRAs) into a unified, coherent, evidence-based argument that the required level of safety has been attained. In addition to furnishing the risk acceptance rationale, the RISC can serve as a roadmap for risk management during deployment, including activities such as precursor analysis, a process for learning from flight experience that is now considered a necessity for risk management of human space flight.

The recently published NASA System Safety Handbook [1] presents the system safety framework and provides guidance for its implementation. It provides a structured model for planning, conducting, and documenting system safety activities in a manner that meets stakeholder objectives and provides the technical basis for risk acceptance. The framework consists of the following system safety elements: safety objectives, system safety activities, risk-informed safety case (RISC) preparation, RISC evaluation, and RISC approval, as illustrated in Figure 1. The framework is implemented in a negotiated fashion between system Acquirers and system Providers. Safety objectives, RISC evaluation, and RISC approval are within the purview of the Acquirer, whereas system safety activities and RISC preparation are within the purview of the Provider.

1 An Acquirer is a NASA organization that tasks a subordinate organization to produce a product or deliver a service. The Acquirer is responsible for safety assurance, i.e., the development of confidence that safety has been sufficiently ensured by the Provider, such that a decision can be made to accept the safety risk of the system. A Provider is a NASA or contractor organization that is responsible for safety assurance, i.e., the reduction and elimination of system hazards and the achievement of adequate safety performance through design, procurement, fabrication, construction, and in the case of a service provider, operation.
Relationship of System Safety to Systems Engineering and Risk Management: The approach to system safety articulated in this Standard recognizes the substantial overlap between systems engineering, risk management, and system safety. Both risk management and systems engineering are concerned with the achievement of organizational objectives. Broadly speaking, systems engineering is the means by which the objectives are met, and the role of risk management is to provide a control function for systems engineering to assure that the development is, and will remain, on track to meet the objectives, across all mission execution domains, including safety, technical, cost, and schedule.

System safety is a subset of both systems engineering and risk management. Considering that the safety performance of a system is a stakeholder concern in much the same way that technical performance capabilities such as payload mass to orbit are stakeholder concerns, system safety is an integral part of systems engineering efforts to develop a system that satisfies stakeholder objectives. System safety is also an objective of risk management efforts to identify and respond to conditions that may arise that threaten the achievement of the system’s baseline safety performance.

Although each mission execution domain has unique characteristics, system performance in the safety domain is of particular concern to NASA, as reflected in safety being a NASA core value [2]. Moreover, safety performance is typically probabilistic, in that adverse safety consequences occur only sporadically, which means that 1) safety performance is not directly observable, but must be inferred; and 2) assessments of safety performance are inherently uncertain, since they are subject to the same margin of error considerations that apply to stochastic phenomena generally.

Relationship of the Evolving Framework to MIL-STD-882: System safety has traditionally focused on identifying and controlling individual hazards, as exemplified by standards such as MIL-STD-882, Department of Defense Standard Practice - System Safety [3], which has been a primary reference document for system safety since its initial release in July of 1969. MIL-STD-882 identifies the Department of Defense (DoD) systems engineering approach to eliminating hazards, where possible, and minimizing risks where those hazards cannot be eliminated. It covers hazards as they apply to systems / products / equipment / infrastructure (including both hardware and software) throughout design, development, test, production, use, and disposal. The intent of the NASA system safety framework is to build upon, rather than replace, standards such as MIL-STD-882 by addressing NASA-specific needs that go beyond those addressed in existing documents.

Safety Objectives and Associated Requirements

As discussed in NPR 8000.4A, Agency Risk Management Procedural Requirements [4], at the outset of a program or project, the objectives, deliverables, performance measures, baseline performance requirements, resources, and schedules that define the task to be performed are negotiated between the organizational unit performing the task (Provider) and the organizational unit responsible for oversight (Acquirer). As part of this process, a set of safety objectives is negotiated consistent with the two fundamental safety principles, namely meeting minimum tolerable levels of safety and being as safe as reasonably practical (ASARP).

Minimum Tolerable Levels of Safety Performance: Setting minimum levels of safety performance involves consideration not only of societal issues relating to safety, but also of what is feasible given current capabilities and technological development potential. It involves the conduct of safety studies by the Acquirer during pre-Formulation to better understand the safety risks involved. Minimum levels of safety performance are not necessarily constant over the
life of a system. An initial level of safety performance may be accepted for a developmental system, with the expectation that it will be improved as failure modes are "wring out" over time. In such cases the level of tolerable safety can be expressed as a safety threshold against which initial system performance is assessed, and a safety goal against which future performance will be assessed. Figure 2 illustrates a safety threshold and safety goal for a generic safety performance measure.

![Safety Threshold and Safety Goal for a Generic Safety Performance Measure](image)

The Aerospace Safety Advisory Panel (ASAP) [5] and others have raised the need to consider the gap between actual risk and explicitly quantified risk when applying safety thresholds and goals. This concern reflects the expectation that during the early stages of operation there is likely to be significant risk from unknown and/or underappreciated (UU) sources. NASA's agency-level safety thresholds and goals do not explicitly address the question of how to account for these contributors to system risk. One possible approach is imbedded in the concept of safety performance margin (referred to in the NASA System Safety Handbook as safety risk reserve). In the safety performance margin approach, the actual risk of a system is understood to be the sum of the risk from known sources, as explicitly quantified using traditional risk analysis methods, plus the risk from UU sources, accounted for by the safety performance margin. This approach is illustrated in Figure 3 in a single-mission context. In a multiple-mission context, or for a single unit operated over a long period of time, the safety performance margin would start at an initial value consistent with historical information about like systems, and subsequently diminish with time as safety performance information is gained through system operation (including tests).

![Safety Performance Margin in a Single-Mission Context](image)

**As Safe As Reasonably Practicable:** A determination that a system is ASARP entails weighing its safety performance against the commitments and tradeoffs needed to further improve it. The system is ASARP if an incremental improvement in safety would require an intolerable or disproportionate deterioration of system performance in other areas. Thus, a system that is ASARP is one where safety is given the highest priority within the constraints of operational effectiveness, time, and cost, throughout all phases of the system life cycle. Being ASARP is a separate and distinct consideration from meeting a minimum tolerable level of safety. ASARP makes no explicit reference to the absolute value of a system's safety performance or the tolerability of that performance. It is strictly concerned with the system's safety performance relative to that of other identified alternatives. ASARP reflects a mindset of continuous
safety improvement regardless of the current level of safety. It is an integral aspect of a good systems engineering process that guides risk-informed decision making throughout the system life cycle. Correspondingly, the condition of ASARP is demonstrated to some extent through process considerations. Figure 4 illustrates the ASARP region for a generic set of alternatives.

Operational Safety Objectives: The two fundamental principles decomposed into specific safety objectives to be met by the system. By specifying safety objectives down to a level where they can be clearly addressed by systems engineering processes, an operational definition of safety is created that enables the processes to be developed and evaluated in terms of the safety objectives. By adequately meeting these so-called operational safety objectives, then by virtue of their derivation from fundamental safety principles, the system can be said to be adequately safe. Figure 5 illustrates the derivation of generic operational safety objectives, including the application of safety performance margin to minimum tolerable levels of safety, so that safety requirements can be developed that maintain a reserve for the expected presence of unknown and/or underappreciated scenarios. The figure shows that even in the absence of minimum tolerable levels of safety, the ASARP principle is still operative. Traditional, deterministic safety practices such as requiring redundancy where practical have implicitly recognized the ASARP principle as fundamental to system safety.

The transition from safety objectives, which are defined by the Acquirer, to system safety activities, which are conducted by the Provider, is accomplished by translating the objectives into safety requirements that the Acquirer then levies on the Provider. The development of safety requirements is the purview of the Technical Requirements Definition Process of the NASA Systems Engineering Engine presented in NPR 7123.1A, NASA Systems Engineering Processes and Requirements [6]. It is carried out collaboratively between the Provider and the Acquirer, but the evaluation regarding the appropriateness of the collection of safety requirements resides with the Acquirer. Early in the system life cycle the Provider conducts a System Safety Requirements Analysis (SSRA). The SSRA serves to clarify what detailed requirements (including, but not limited to, engineering requirements) the Provider expects to address in the ensuing development, and which form the basis of the Provider’s System Safety Management Plan (SSMP). In addition, the SSMP focuses on the collection of process requirements. Figure 6 illustrates the interaction/iteration between the Acquirer and Provider during the development of system-specific safety requirements.
Figure 5 - Objectives-Driven Requirements Development (notional)

Figure 6 - Interaction/Iteration between the Acquirer and Provider during the Development of System-Specific Safety Requirements
System Safety Activities

System safety activities are conducted by the Provider as part of overall systems engineering technical process activities. System safety activities are focused on the achievement of the stated safety objectives. System safety activities not only ensure the safety of the system, but also produce the evidence of safety that will be used to support claims in the RISCs provided to the Acquirer at key decision points in the system life cycle. Because of the diversity of Providers, each with its own particular set of systems engineering, system safety, and risk management processes, NASA recognizes the need for flexibility in the nature and composition of system safety activities, as long as they are able to achieve their operational safety objectives. System safety activities typically fall into the general categories of: Conducting an Integrated Safety Analysis (ISA); Requirements Development Support; System Design Support; Program Control and Commitments Support; and Performance Monitoring Support.

Conducting an ISA: ISA includes methods for identifying and characterizing potential accident scenarios. This includes accident causes, contributing factors, effectiveness of controls (both existing and proposed), subsystem interactions, analysis of physical responses of the system to the environments it encounters, and analysis of the probability that the undesirable consequences will be realized. ISA integrates different types of safety analyses (e.g., FMEA, PRA, phenomenological modeling) to the greatest extent possible. The ISA consolidates these separate analyses to produce a single comprehensive set of quantified safety performance measures that can be used to assess the standing of the system with respect to the levied safety performance requirements. The ISA is then used to risk-inform system design and operational decision-making. The ISA must be tailored to the particular phase in the life cycle at which it is conducted. As the system design evolves, the ISA is kept current, typically through the use of progressively more rigorous analysis techniques that model the system at progressively finer levels of detail. The ISA is maintained during system realization so that it can be used to inform decisions related to safety, such as test protocols. During system operation, the ISA is updated to reflect such things as design modifications and accumulating operational experience, including anomalies.

Requirements Development Support: Through the early conduct of an SSRA, the Provider identifies applicable requirements by reviewing NASA, military, and industry standards and specifications, historical documentation on similar and legacy systems, etc. Additionally, using the ISA, the Provider translates any system-level probabilistic safety performance requirements levied by the Acquirer into objectively verifiable system-specific derived requirements (and associated verification procedures) that protect the assumptions underlying the system’s assessed safety performance. Such derived requirements may involve, for example, prescribing specific levels of component reliability, specifying limits on environments produced by components such as particulate emissions or vibrations, or requiring a certain level of failure tolerance in a subsystem. In cases where the flowdown of requirements crosses organizational boundaries, the ISA provides a rational basis for allocating requirements to subordinate organizations. It may be the case that a levied requirement proves to be overly burdensome (such as by adding too much mass to the system) or sub-optimal (e.g., where alternate means are available to meet the intent of the requirement). System safety plays a role in these cases by assessing the potential consequences of tailoring the requirement, both through explicit modeling of safety performance using the ISA, and by qualitative consideration of the potential erosion of protection against unknown and underappreciated scenarios. This provides a technical basis for tailoring the requirement.

System Design Support: System design support is of two broad types: best-practice-informed and ISA-informed. Best-practice-informed design support promotes safety by identifying applicable historically-applied safety-related engineering requirements and by assuring that proven strategies for optimizing safety are considered during system design decisions. ISA-informed support promotes safety by risk-informing design decisions with an assessment of the safety performance of each contending alternative. These two types of design support work synergistically to achieve a design that is ASARP.

Program Control and Commitments Support: System safety promotes the development of program controls and commitments needed to ensure that the framework for safety is backed by sound administrative and management practices. Of particular importance to the maintenance of the system’s safety performance is the identification of safety-critical items (SCIs) that are explicitly relied on for safety. A major vehicle for SCI identification is the ISA, which is used to identify the hardware, software, human, operational, and managerial system features upon which safe system operation depends. Such items can be explicit in the ISA (e.g., redundancies, backup systems) or they can be implicit (e.g., assumptions regarding component structural integrity). In either case, designating these items as safety-critical protects their safety functions by imposing rigorous and highly visible safety management provisions on them. The adequacy of safety-critical item designation is ultimately at the discretion of the Acquirer, and may include (for example) items associated with safety-related engineering requirements levied for reasons such as defense-in-depth or margin preservation, independent of the ISA. Other aspects of program controls and commitments covered within the system safety framework but not necessarily as part of an ISA include configuration management, quality assurance, training and certification of personnel, use of best practices and lessons learned, and assurance that safety requirements are being complied with. Any specific levied requirements, controls, and commitments that are considered critical to safety could be designated as SCIs.
Performance Monitoring Support: System safety supports effective performance monitoring, both in the development of monitoring protocols and in responding to performance data. The ISA is used to risk-inform the selection of system attributes that will be monitored, both to ensure that significant uncertainties are reduced as experience accumulates and to ensure that important performance-related assumptions in the ISA remain valid over the system life cycle. Anomalous performance data are scrutinized for their potential impact on safety (e.g., via accident precursor analysis [7]) and managed accordingly.

RISC Development

A RISC is a structured argument, supported by a body of evidence that provides a compelling, comprehensible and valid case that a system is or will be adequately safe for a given application in a given environment. It addresses each of the operational safety objectives of the system, including plans for achieving safety objectives that are applicable to later phases of the system life cycle. In the context of NASA systems engineering, the RISC refers to the totality of safety-related documentation submitted to a given technical review. As such, the documentation requirements of the RISC are consistent with the entrance criteria for the relevant review, as itemized in NPR 7123.1A. Similarly, the criteria for evaluating the adequacy of the RISC are consistent with the corresponding technical review success criteria in the same NPR.

The elements of the RISC are [8]:

- An explicit set of safety claims about the system(s), for example, the probability of an accident or a group of accidents is low
- Evidence justifying the claims, for example, representative operating history, redundancy in design, or results of analysis
- Structured safety arguments that link claims to evidence and that use logically valid rules of inference

The interaction of these elements is illustrated in Figure 7 for a safety claim supported by two independent arguments.

The concept of adequate safety requires that safety be addressed throughout all phases of the system life cycle. Correspondingly, the RISC must also address the full system life cycle, regardless of the particular point in the life cycle at which the RISC is developed. This manifests in the RISC as two distinct types of safety claims:

- Claims related to the safety objectives of the current or previous phases argue that the objectives have been met.
- Claims related to the safety objectives of future phases argue that a ‘roadmap’ has been established for the satisfaction of objectives yet to be met, i.e., that necessary plans, preparations, and commitments are in place to meet safety objectives at the appropriate time.

The form of the RISC arguments for accomplished objectives vs. upcoming objectives is shown in Figure 8 for the point in time at which design has completed and realization is about to commence. As the system proceeds in the life cycle and RISCs are developed for successive milestones, arguments demonstrating an ability and commitment to meeting objectives are replaced by arguments demonstrating accomplishment of objectives.

Upon submittal of the RISC to the Acquirer, the Acquirer, usually through a designated Evaluation Team (referred to as the Evaluator), conducts an evaluation of the RISC to determine the technical adequacy of its safety claims. RISC evaluation is carried out based on defined evaluation protocols for the system at a particular point in its life cycle. For each claim in the RISC, it is the task of the Acquirer to:

- Understand the evidence behind the claims
- Evaluate the evidence to determine its validity
- Provide judgment as to validity of the claims

In other words, to evaluate the RISC, the claims in the RISC are critically reviewed, thereby making use of the collected evidence related to the safety of the system.

![Diagram of RISC evaluation process]

**Figure 8 – Coverage of the System Life Cycle in the RISC**

**RISC Evaluation**

The Evaluator ultimately rates the RISC overall as being Acceptable or Unacceptable. In order to provide rationale that the RISC is acceptable, the Evaluator must be able to infer from the evidence in its totality that the top claim (e.g., that the system is adequately safe) has been substantiated to high confidence. The ability to make this inference is based on the Evaluator's knowledge of the system as a whole and of the various combinations of requirements that have to be satisfied in order for the system to be deemed safe. In the evaluation process, it is important for the reviewer(s) to evaluate the RISC from a critical viewpoint, examining the supporting evidence as necessary to develop confidence in the claims at all levels. The output of the RISC evaluation is a set of evaluation findings summarizing the review and indicating potential areas of weakness in the RISC.

The expectation is that the prospective Provider should understand in advance the eventual evaluation process to which the RISC will be subjected. Safety case evaluation is part of the Acquirer’s risk management process. It is a particular instance of supporting a risk acceptance decision. Requirements of NPR 8000.4A must therefore be addressed. In particular, the risk communication protocols established between the acquiring organization and the providing organization should be understood to be operating within the framework set forth in NPR 8000.4A.

**RISC Approval**

RISC approval is granted when the Acquirer is satisfied that the technical basis of the RISC is sound. The approved RISC, communicated in the form of a RISC Report, and the RISC evaluation, communicated in the form of a RISC Evaluation Report, become the safety-specific technical bases supporting the decision for which the RISC was developed. RISC approval entails a commitment to maintaining the validity of the RISC as approved, upon which the Acquirer’s decision will be predicated. As such, the commitments and understandings captured in the RISC approval become part of the performance baseline to be managed subsequently under the Continuous Risk Management (CRM) elements of NPR 8000.4A. The risk of a shortfall in safety performance relative to this baseline is managed under the same risk management process within which all other performance risks are managed. If thresholds have been established for safety performance risk, then risk elevation will be required should threshold satisfaction be threatened by emergent conditions. NPR 8000.4A imposes requirements on organizational units (e.g., Acquirer and Provider) to coordinate risk management between the two organizations. The requirements in NPR 8000.4A are general, cutting across mission execution domains, but, as noted previously, the high-level protocols in this Handbook are essentially system-safety-specific instantiations of the general risk management requirements in NPR 8000.4A.

**System Safety throughout the System Life Cycle**

NASA programs and projects are managed to life cycles, the division of the program’s and project’s activities over the full lifetime of the program or project, based on the expected maturity of program and project information and products as they move through defined phases in the life cycle. Figure 9 shows a simplified version of a project’s life cycle to illustrate the relationship between the phases, the key decision points and application of the system safety framework. (Program and project life-cycle phases are described in NPR 7123.1B.) The vertical thickness of each shape in the
The NASA System Safety Handbook presents a system safety framework that provides a coherent structure for organizing system safety activities towards the achievement and demonstration of safety throughout the system life cycle. Within the framework, system safety activities are organized around the accomplishment of clearly stated safety objectives that collectively define adequate safety for the system, and are communicated to decision makers via the RISC, which provides a compelling, comprehensible and valid argument, supported by evidence, that the system is or will be adequately safe for a given application in a given environment.

Acknowledgements

The authors express their gratitude to present and past NASA Office of Safety and Mission Assurance (OSMA) leaders for their support and encouragement in developing this document. The authors also recognize the contributions made by Bob Youngblood (INL), Allan Benjamin (consultant), Curtis Smith (INL), Dev Sen (ISL), Martin Feather (JPL), and Pete Rutledge (QA&RMS). The development effort leading to this document was conducted in stages, and was supported through reviews and discussions by the NASA System Safety Steering Group (S3G) and others.

References


Biography

Homayoon Dezfuli, Ph.D., NASA Headquarters Office of Safety and Mission Assurance, 300 E Street SW, Washington, DC 20546, email - hdezfuli@nasa.gov.

Dr. Dezfuli is the NASA System Safety Technical Fellow and the Manager of System Safety at NASA Headquarters OSMA. In these roles, he serves as a senior technical expert for the Agency in system safety methodology and practice and leads NASA's policy development initiatives for system safety and risk management. Dr. Dezfuli has been instrumental in developing and implementing advanced system safety and risk management techniques and processes for the Agency. He led the development of and co-authored the NASA Probabilistic Risk Assessment (PRA) Procedures Guide, NASA Risk-Informed Decision Making Handbook, NASA Risk Management Handbook, and NASA System Safety Handbook. He is also the co-author of the NASA Systems Engineering Handbook and is the author or co-author of many papers in the areas of safety, risk assessment, and risk management. He devised a safety goal implementation framework that has helped shape the NASA safety goal policy for human space flight. Dr. Dezfuli is currently leading the development of a case-based assurance framework for Safety and Mission Success. Dr. Dezfuli has a Ph.D. in nuclear engineering from the University of Maryland.

Chris Everett, Energy and Space Division, Information Systems Laboratories, 710 West End Ave., New York, NY 10025, email - ceverett@islinc.com.

Mr. Everett is a safety and risk analyst with twenty-eight years experience supporting government agencies and private-sector clients in the areas of systems safety, risk assessment, and risk management. He is currently supporting NASA in the development of system safety and risk management processes and guidance, the conduct of risk assessments, and the development of computer-based risk assessment tools. He is a co-author of the NASA Accident Precursor Analysis Handbook, the NASA Risk-Informed Decision Making Handbook, the NASA Risk Management Handbook, and the NASA System Safety Handbook. He was project manager and co-developer of the Dynamic Abort Risk Evaluator (DARE), which assessed the effectiveness of Shuttle and Ares I aborts.

Frank Groen, Ph.D., NASA Headquarters Office of Safety and Mission Assurance, 300 E Street SW, Washington, DC 20546, email - frank.j.groen@nasa.gov.

Dr. Groen is the Director of Safety and Assurance Requirements Division (SARD) within NASA Headquarters OSMA. In this position, Dr. Groen is responsible for the development and maintenance of NASA directives and standards pertaining to Safety and Mission Assurance, as well as related methods, tools and guidance. Prior to his current position, Dr. Groen worked in OSMA as the Manager for Reliability and Maintainability at OSMA, and also served as the document manager for NASA's Human Rating directive and program executive for NASA's ELV Payload Safety program. During this period, he introduced the safety goal policy for human spaceflight missions into the Human-Rating directive, oversaw the development of an accident precursor analysis methodology for NASA, and initiated an objectives-driven approach for the standardization of R&M activities. Prior to entering NASA, Dr. Groen was active in academia and industry, where he focused on method and tool development in the field of reliability and risk assessment, with a focus on Bayesian data analysis, accident scenario modelling and analysis, and Monte Carlo simulation. Dr. Groen received his PhD in Reliability Engineering from the University of Maryland in 2000 and a MSc in Mechanical Engineering from the Delft University of Technology, The Netherlands.