CSER 2017 conference on systems engineering research

15th Annual Conference on Systems Engineering Research
Disciplinary Convergence: Implications for Systems Engineering Research

Eds.: Azad M. Madni, Barry Boehm
Daniel A. Erwin, Roger Ghanem; University of Southern California
Marilee J. Wheaton, The Aerospace Corporation
Redondo Beach, CA, March 23-25, 2017

System safety data network: Architecture and Blueprint

Shravan Shetty, Dr. Mark Avnet, Dr. Farzan Sasangohar
Texas A&M University, shravanshettyu@tamu.edu
Texas A&M University, avnet@tamu.edu

Abstract

With increasing complexity of safety analysis in socio-technical systems, there is a need for a mechanism to accurately capture complex information and present it in an easily accessible and understandable form. While there are plenty of accident databases that have been created over the years for specific purposes, a tool that provides a homogeneous view of all the safety-related aspects of an accident customized specifically per user and industry is largely absent. This paper discusses the conceptual model of The System Safety Database (SSD), a tool that will offer tailored solutions to multiple classes of users and that will generate reports synthesizing lessons learned from a variety of disparate contexts, providing succinct and actionable information for decision support. Here we also propose the concept and architecture of a Safety Data Network (SSDN) that encapsulates a network of safety databases, thereby addressing some of the challenges of a stand-alone safety database. The data network will enable working with structured and unstructured data by integrating multiple Relational and NoSQL databases. A full-fledged implementation of the safety data network will enable improved collaboration across industries and corporations. The safety data network will facilitate analysis across disciplines and contexts, allowing researchers and practitioners to use integrated mixed-methods approaches to conduct investigations, analyses, research, and development activities across multiple levels of a system. The paper also discusses the steps involved in the implementation of such a data network and the challenges involved. In addition, the current work in data categorization and interpretability of incident data is discussed. When completed, the System Safety Data Network will provide stakeholders at all levels, from individual operators to policymakers, with the tools and perspectives needed to improve the safety of complex socio-technical systems.

Keywords: System safety, system safety data network, system architecture, accident investigation, accident case studies

1. Background and motivation

With advancements in accident investigation models [e.g., 1] and the introduction of analytical methods in accident investigations, it is now generally accepted that accidents in sociotechnical systems are rarely due to single isolated causes [2]. To illustrate this point, let us consider the Swissair Flight 111 accident. Though the reported cause was a faulty wire that initiated the fire, a convolution of lack of clear regulations regarding in-flight fire control, poor crew training, and highly flammable thermal insulation blankets led to the fire engulfing the cockpit [3]. This emphasizes the importance of interrelationships between sub-events and subsystems, and the necessity to capture and analyze such complex information to develop a holistic understanding of accidents. The value and learnings from such information would arguably increase if it was easily available across domains and industries in a single central repository.

The concept of a safety database is by no means novel. Many organizations, such as the Aviation Safety Network (ASN) and the National Aeronautics and Space Administration (NASA) host databases that contain

reports on thousands of accidents [4, 5]. But the value drawn from these reports are usually limited at best, as a systematic and in-depth analysis of accidents is scarcely conducted and the data available is rarely easily accessible. There is no dedicated tool to analyze and integrate the learnings from available data and apply its knowledge to benefit safety across all industries.

In our previous work, we conceptualized the System Safety Database as a universal repository of information about accidents, regulations and regulatory bodies, expert analyses, and safety methods and frameworks across various industries [6, 7]. Systematic analysis of cases using the Multilevel Frameworks [8] and case-based reasoning approach [9] was conducted on 7 individual cases. During implementation, it was identified that the challenges to a universal repository of accidents was a two-phased problem, with unravelling the depth and complexity of information being one and the issue of creating a collaborative environment for sharing safety information being the other. One of the most important limitations of the suggested model was that safety information or accident data is not easily available, especially with incidents related to government organizations or corporate entities and a system providing incentives and encouraging collaboration needed to be developed.

The goal of this paper is to describe the architecture and blueprint of a System Safety Data Network (SSDN) and to provide an update on the implementation efforts. A multidisciplinary approach involving a team comprising of members from nuclear, petroleum, aerospace engineering, computer science and industrial engineering domains is being used for the development of the network. The SSDN is designed to a be a central, service-based data network that in addition to the features of the SSD, provides the ability of secure and controlled data collaboration for third-party collaborating organizations. The internal databases will host systematically analyzed information from a broad spectrum of industries and will be able to generate customized reports for particular stakeholder groups and system contexts. In addition, the Application Program Interface (API) services provide collaborating organizations with the ability to incorporate analytical and data services into their personal systems while providing methods to contribute controlled, anonymized and secure content to the safety data network at their consent. The paper also describes the challenges and opportunities provided by a full-fledged safety data network.

2. Methodology

In order to effectively capture the data to build a repository of accident information, a comprehensive semi-structured interviews with 16 industry experts with expertise in safety across a variety of application domains and disciplines was conducted. Participants were sampled using maximum variation (Patton, 1990) based on their areas of expertise to ensure a broad base of information. Interviews ranged in length from 30 to 90 minutes and were conducted in person or over the telephone. The interviews were also audio-recorded and transcribed. All interviews were based on a common set of questions in a pre-constructed interview guide. The interview guide contained questions on the participants' background and their relative knowledge and experience with safety incidents, and also open-ended questions on any summary thoughts or perspectives that they would like to add. Further context-based probes were used to gain additional information, clarification, or expansion as needed. To transcribe the audio files, transcribers listened to each audio transcript and typed the data into a file. A back-referencing technique was used until each file was fully accurate. Upon completion of the interviews, results were analyzed and a list of relevant attributes was developed. The data collection was then split into two steps involving case analyses. Initially, to gather a wide range of data for the repository, a breadth-first approach was used on 40 cases. The cases were carefully chosen as to cover a variety of domains and industries. Next, 7 cases were selected from this set for in-depth analysis based on the content available in public domain. A method for in-depth analysis to pick relevant information from cases was developed using the following three requirements:

- Each case analysis must contain enough information that a random, uninformed user could grasp what caused an accident occurred within a reasonable level of understanding.
- While analysis must be fully realized, each analysis must also be relatively compact in size. Given the full scope of the database and the volume of data included in the final version, the contents of each analysis must be as short as possible to conserve database memory. In practice, this means accident summaries are generally limited to a paragraph.
- Each analysis should strive to use primary sources when possible. Accident reports from regulatory

organizations are preferable. As opposed to primary sources, secondary sources often limit the scope of an accident and do not include the depth required to pick out underlying causes. Also, secondary sources are not always archived, meaning that a source cited in the databased might be impossible to find after a given period of time, leaving entries in the database unverified.

Next, an in-depth analysis of the collected cases was performed by in-house researchers. The underlying causes for an accident were identified, and each underlying cause was listed and classified as being on a technical, human, organizational, or societal level. Dissecting cases to root out underlying causes is a time consuming process, so efforts were directed at fully analyzing cases that expand upon the information derived from other cases. From the data gathered by systematic analysis of 7 cases, an Entity Relationship diagram of the internal database was created. It was then modified to build on the important information specific to individual cases. The class diagrams where then developed and individual attributes were mapped to the data architecture. An alpha version of the database was built and data was loaded into the database. Test queries were conducted on the database to insure the data retrieved was of expected format and quality. Reviews and feedbacks were collected from experts. Drawbacks of previous attempts at safety databases were studied. Based on analysis of feedbacks and lessons from previous databases, the idea of distributed architecture was proposed. After analyzing the technical feasibilities, advantages and disadvantages, an alpha version of the safety data network was developed.

3. System Architecture

One of the primary concerns with a collaborative central repository of safety information based on previous attempts [4, 5] was the reluctance of corporations or government organizations to share accident information due to the secure nature of the information, or in some cases the tendency to hide past mistakes. Hence, the Architecture of the SSDN was developed focusing on data security, collaboration feasibility and data diversity as primary factors. The design consists of a central core that is in the public domain containing information analyzed by in-house researchers and those made available by willing private organizations. To promote sustenance and accuracy of information in the system, the feature for data insertion from individual users with valid proofs is envisioned. The design intends to motivate whistle blowers and anonymous users, making the most accurate information available in the public domain. Individual organizations could then develop their exclusive nodes and interfaces. The participating organization will have access to all the information available in the public core along with their own proprietary node and when possible and made available, access sanitized information from other organizations' proprietary nodes. The organization be in control of data pertaining to its proprietary node and overtime, can work with in-house researchers to sanitize and anonymize information to be shared to the public core.

The SSDN's front-end can benefit from a web service providing users a rich and responsive experience. The data will be provided dynamically by a backend Communication Management Interface (CMI) which is the central data management component of the system. The unit also manage secure logins into the system by looking up into an encrypted login database containing user information. The CMI is also connected to a series of internal and external API's through which content request and responses are conducted. The API's are connected to multiple databases management systems through which the requested information is processed. Here the requests are broken down into queries and data is retrieved from the databases. The databases themselves are distributed comprising of a mixture of relational and NoSQL databases. The relational part captures the structured, surface-level information and also defines the relations between data components. The NoSQL databases capture all the unstructured and complex information in its entirety. This architecture enables the storage of both structured and unstructured form of data.

To illustrate how the System Safety Data Network can be used, we use a recent accident: the disappearance of Malaysia Airlines Flight MH370 [10]. On March 8, 2014, flight 370 from Kuala Lumpur to the Chinese capital, Beijing, lost contact with air-traffic controllers and since then several investigations into the accident have taken place. Using the SSDN (Fig. 1), a user investigating the accident would make a request on information available through the SSD Front end (Section 1). The CMI/LV receives this request, verifies the user and looks up the data dictionary for the location of the relevant data (Section 2). The data might in this instance be a culmination of information from our internal database, the department of civil

aviation Malaysia's database, Federal Aviation Administration (FAA) database, and NASA's Aviation Safety Reporting System Database. Based on the information in the data dictionary, the CMI/LV makes API requests to relevant systems (Section 3 and Section 5). The requests are processed by the systems independently as per its business logic and relevant information is returned in a predetermined response format. This data is loaded back to the front-end dynamically completing one request cycle.

The service-based nature of the architecture has the following strengths:

- Data available to the user will be in the most up-to-date form of information as the information is obtained at runtime via API requests from multiple relevant sources as opposed to having pre-scheduled batch jobs pulling in information into internal databases. For example, if information on the Malaysian Airlines case is updated in the NASA database, traditionally this data will not be reflected to user's in a central repository until it is updated in the internal database, but with a service-based design, the data updated in the NASA database is immediately available for users in SSDN.
- Data security, ownership and availability will be controlled by third-party partners, improving trust in the collaborative effort and also decreasing liabilities on sensitive data. A service-based architecture provides control of data to the collaborating partners, decreasing security concerns while providing a stepping stone for future sharing of anonymized data to the internal databases. Such an architecture would specifically improve the possibility of collaboration by corporate or government organizations. The modular nature of the system enables the system to have high availability. Since a request will involve calls to multiple sources, the probability of all the sources becoming unavailable simultaneously would be very low.

Such distributed architecture however has several important limitations:

- There is limited control on responsiveness of the system since the response of the systems depends on responses of collaborating systems and may suffer as request load increases. Also, with the growth of the data network and the addition of new systems, responsiveness may become sluggish.
- Though there is no single point of failure, full availability of the system will be affected if one of the collaborating systems become unavailable, data specific to that partner becomes non-accessible.
- Also, because partners are responsible for system maintenance, partners' willingness to engage dedicated resources for the maintenance of such collaborative system is uncertain.

In the next section, we discuss the blueprint for a successful SSDN and propose a method to visualize dependencies between its various phases.

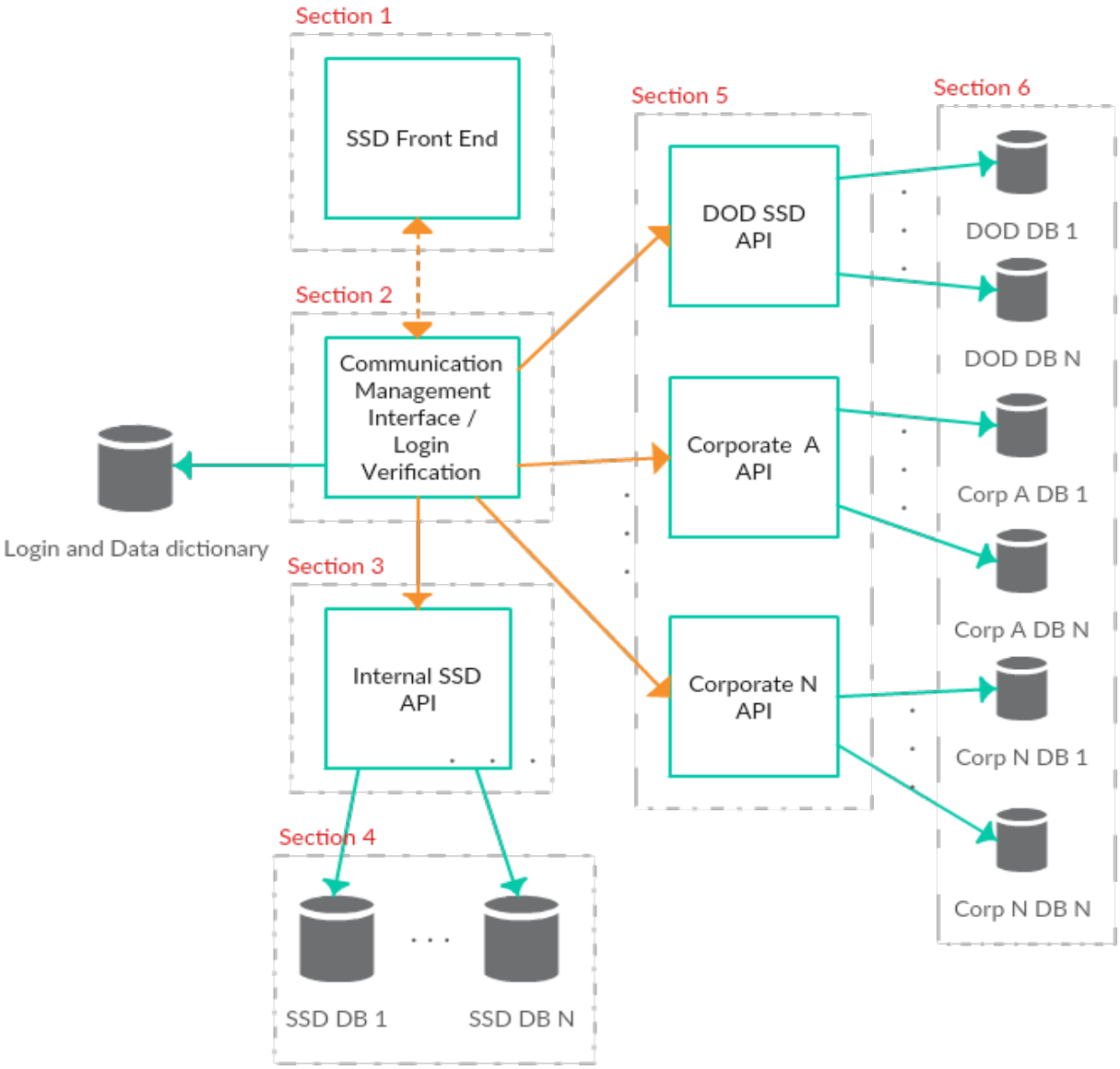

Fig. 1. High level visualization of SSDN's modular Architecture

4. Blueprint for a System Safety Data Network

The system safety data network is still at its infancy and considering the scope of the project, there are a lot of variables to be considered. In this section, we provide a high-level methodology for the implementation of a full-fledged safety data network (Fig. 2). We discuss the scope of the work in progress and various intermediate objectives the team is working towards.

Step 1: Analyze and decompose the accident cases available in the public domain and building an internal collection of accident information. This process is implemented by in-house researchers and the data extracted by analyzing accident cases is discussed in the System Safety Database: Use cases and applications [7]. The basic data catalogued has the following structure:

- The date of the accident

- The number of fatalities or injuries
- Monetary damages (if information is available)
- The proximate cause of the accident
- The duration of the accident
- The responsible organization(s)
- The industry of the organization involved in the accident
- Any regulatory organizations entrusted with safeguarding the responsible organization(s)
- Underlying causes for an accident
- Classification of the underlying causes

Step 2: Build internal database architecture templates to accurately capture the diverse structured and unstructured data captured during the analysis and extraction phase. This phase involves evaluating the structure of the available data, analyzing the various complex interrelations between these data structures and building an architecture that accommodates and integrates these relationships.

Step 3: Build an alpha version of the internal databases on a local server and implement the integration of the diverse databases. This phase could benefit from visual representation of complex accident information using the concept of network theory. In a recent effort the sources of perceived complexity and their relationships were captured in a visual format to help users understand the complex nature of several nuclear accidents and to easily map the interrelations and major factors [11].

Step 4: Identify user groups and build a core group of prospective users to understand the needs of each user group and map requirements in collaboration with these prospective users. This phase involves market research by conducting interviews, surveys and polls to build a comprehensive set of requirements for each group of users interacting with the tool. Based on the analysis of these requirements, use cases and activity charts are developed to map the functionality of the system.

Step 5: Build the backend architecture based on the research done in step 4 and wrap the internal functionality in modules and expose the required interface methods using API's. This step involves building the business logic of the system using a modular approach and integrating the internal API modules to backend databases. Also, this step involves documenting and building a demo third-party interaction API, which will act as a reference and provides expected communication protocols to the data collaboration partners. We acknowledge that the implementation of the partner API's will have to be customized on a case-by-case basis depending on the architecture and structure of the collaborating partner systems.

Step 6: Create a business team in charge of growth of the data network, by approaching corporations and government organizations for data collaboration using API's. This will be a crucial phase in the development of the database as convincing reluctant parties to share data is one of the biggest challenges that has hindered the development of a centralized data repository.

Step 7: Build a responsive, maintainable and dynamic frontend interface for the SSDN that interacts with the backend via API's. A modular design will help with incremental addition of features to the system, thereby providing flexibility for upgrades.

Step 8: Beta-test the system with a core group of trial users, collecting data on usability and interactivity of the systems features. A feedback loop and an active user community will sustain the continuous development of the tool.

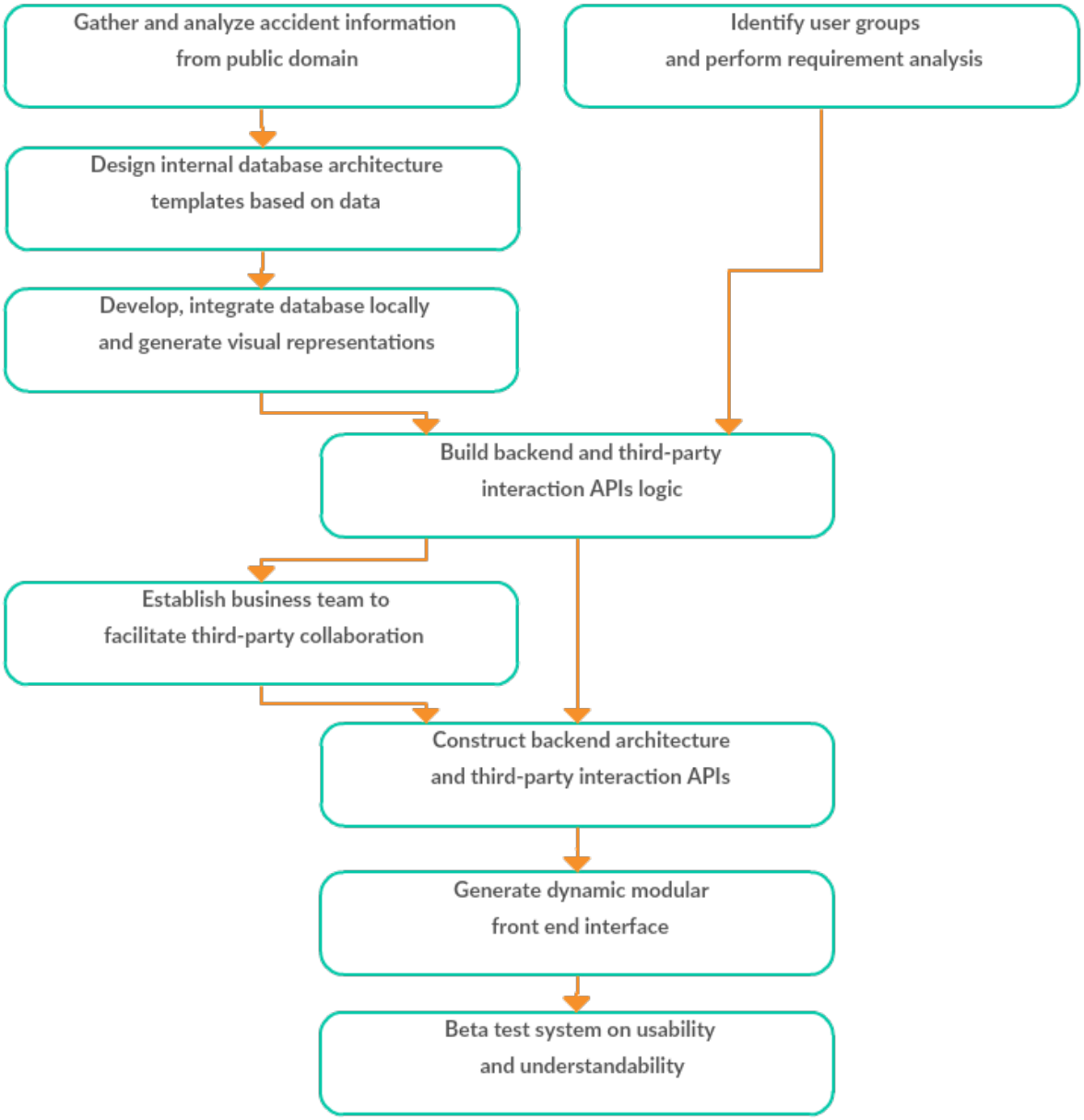

Fig. 2. Blueprint of phases involved in fulfilment of the SSDN

The above-mentioned blueprint for implementation of a safety data network would set the stage for extensive data analysis on accidents and develop a deep understanding of accident complexities across industries.

5. Current Work-in-Progress

The work on data extraction and analysis started in Fall 2015 and a breadth-first analysis of 40 accidents as well as in-depth systematic analysis of 7 incidents has been completed. The database architecture capturing the surface-level information was completed in December 2015 and currently work is being conducted in mapping the complex interrelation between causes and capturing the holistic view of the accidents. The team started exploring NoSQL databases in May 2016 and are currently looking into suitable

options for the application. An alpha version of the relational database was built on MySQL and data from the 22 analyzed accidents were loaded. A preliminary market research on the user groups was conducted and based on the requirements gathered from the research, basic use cases for the system was developed and presented at the 2016 Industrial and Systems Engineering Research Conference (ISERC) which concluded in May 2016. After consulting industry experts and analyzing the challenges in the development of the system safety database, the idea of a data network was proposed. The current architecture was decided upon after analyzing the strengths and weaknesses of all the proposed architectures and conducting a suitability analysis mapping requirements against the strengths and weaknesses of models.

6. Conclusion: Challenges and Opportunities for the System Safety Data network

In this paper, the need for a comprehensive collaborative central repository containing information on the underlying causes of accidents, regulations, safety analysis methods, and experts was discussed. The system would be open source and would also accept verified contributions from anyone who wishes to add information to the system. It would also contain collaboration partners who would contribute to the repository by opening screened and anonymized sections of their internal data to the network via API services. The paper further discusses the architecture of such a data network and talks about the blueprint to building the database. It also provides an update on the current work, status and challenges of the project. The applications of a fully functional data network include understanding of risk factors in system, enumeration of relevant regulations, collaborative work space for research and many more. Once completed, the System Safety Data Network will provide analytical tools and generate tailored reports for all stakeholders to measure and make data-driven decisions, thereby improving safety of complex engineered systems.

Considering the scope and complexity of the project, white space risks need to be accounted for and the blueprint needs to be flexible to incorporate changes due to requirements or challenges that occur during implementation. A thorough analysis of strengths, weaknesses, opportunities and threats (SWOT) needs to be conducted and a business plan for the system needs to be developed. The current approach to take the initial steps in analyzing databases and collecting data from specific industries and building on top of the in-house knowledge of the incidents. The bulk of the future work involves addressing the challenge of mapping the complex inter-relations to data structures on a case-by-case basis. This includes unifying and mapping accident, regulations, system analysis methods, and export information, as well as creating a method for generating basic safety checklists using the available information. Comprehensive and robust definitions of complexity, scale, and scope of accidents need to be standardized and visualization mechanisms to incorporate more complex information in easy to understand manner needs to be strategized.

Sustenance and maintenance of the network over a period of time is achieved by maintaining the public core as open source and commercialization of the private nodes. Open source of the central repository promotes an active user base committed to maintaining safety information accessible and accurate. To achieve sustenance, use cases on user data insertion and automated validations through dynamic channels need to be incorporated into the design of the system. Allowing users the ability to insert or update case information provides an alternate channel of undocumented information on the incident into the system. The design of the system should empower and motivate users to be the flag bearers of an open, accurate, and collaborative safety information data network.

With respect to maintenance of the system, upon completion of the fundamental research and development of the core public node, the collaborating organizations build and maintain their own proprietary nodes with reference from the internal team managing the demo third-party interaction API. For the maintenance and update of the core central components, a non-profit organization or board would be required which would have exclusive responsibility to keep the information accurate and accessible. A systematic survey of available organizations with the right motivation, vision and commitment to accessible safety information needs to be conducted in this regard.

Another aspect to be considered in the creation of a central database implementation is the acquisition of the data. Though much of the data in the civil sector (in the United States) has been made available through the Freedom of Information Act (FOIA), the bureaucracy involved in submitting and responding to a FOIA request presents a challenge. In the defense sector, much of the data are classified and often cannot be shared externally to the Department of Defense (DoD). Finally, in the private sector, companies

often are incentivized to respond to regulations by ignoring or hiding information about safety incidents rather than using that information to reduce the chances of future accidents. For this reason, these companies are often unwilling to divulge the data needed for a repository such as the SSDN. Hence, Convincing prospective partners about the security and advantages of a data network and emphasizing the issues of control on data collaboration information is critical to the success of the project. The acceptance and of approval of partners to actively collaborate via the SSDN will go a long way in the creation of a central repository of safety information.

Acknowledgements

The authors would like to thank the 16 expert researchers and practitioners that leant their time and expertise to participate in interviews for a related study and whose ideas contributed to the foundation for the database. Early development of this research was supported by the Systems Engineering Research Centre (SERC) under an RT 128 research incubator grant from the Office of the Deputy Assistant Secretary of Defense for Systems Engineering (ODASD(SE)) project. The authors would also like to acknowledge Nanditha Soundararaj (MIS, Texas A&M University) for her input and support in finalizing the paper.

References

1. Oakley J. *Accident investigation techniques*. 2nd ed. American Society of Safety Engineers; 2012.
2. Public Education and Conference Section, *Fixing the System with Root Cause Analysis*. Oregon Occupational Safety and Health Division.
3. McDonnell, D., Swissair Transportation Limited, 1998, *In-Flight Fire Leading to Collision with Water*. McDonnell Douglas MD-11 HB-IWF Peggy's Cove, Nova Scotia 5 nm SW.
4. Aviation Safety Reporting System Database. NASA. Accessed Nov. 8, 2015.
5. ASN Aviation Safety Database. Aviation Safety Network. Accessed Nov. 8, 2015.
6. System Safety Database: Challenges and Opportunities, CESUN conference 2016
7. *System Safety Database: Use cases and Applications*, ISERC conference 2016
8. Avnet, S.M., and Smith-Jackson, L.T., 2015, *A Multilevel Framework of System Safety: Technical Failures, Human Factors, Organizational Behavior, and Societal Influence*, T.A.M. University, Editor: College Station, Texas (in White Paper).
9. Bergmann, R., Althoff, K.D., Minor, M., Reichle, M., Bach, K., 2009, *Case-Based Reasoning – Introduction and Recent Developments*, German Research Foundation
10. *Malaysia Airlines Struggles to Salvage Its Image a Year After Flight 370 Disappearance*, in Time Magazine. 201
11. Sasangohar, F., *A Holistic Investigation of Complexity Sources in Nuclear Power Plant Control Rooms*, Masters Thesis, MIT.