Proceedings of the 2016 Joint Rail Conference
JRC2016
April 12-15, 2016, Columbia, SC, USA

JRC2016-5799

REVIEWING THE USE OF PROACTIVE DATA ANALYSIS IN DEVELOPING RAIL SAFETY CULTURE

Greg Placencia, Ph.D.
University of Southern California
Los Angeles, California, USA

ABSTRACT

While experience is often the best teacher, learning from precursors is much less painful. The aviation and health care industries have greatly benefited from proactively analyzing and developing measures to address sentinel events and learning from various data sources. Such reflective learning is typical of High Reliability Organizations (HROs) with strong learning cultures. As technology like Positive Train Control increasingly integrates into the rail industry, the resulting data they inevitably produce can provide a wealth of knowledge that can greatly improve safety if the data streams are well managed and not blindly mined. For example, simulators generate data while locomotive engineers use them. During training, such data can indicate weak points where the engineer can improve. Examining such data over multiple engineers can establish general areas of strengths and weaknesses among trainees where instructors can place more or less focus and develop better overall training options. Such data could potentially be used to improve cab design and establish how trains and cab care would operate along a given rail line. This paper will explore the use of data streams from various sources, including those currently used like injury reports, emerging ones like simulation training evaluations and data logs to develop better safety cultures within the rail industry.

INTRODUCTION

Positive Train Control (PTC) generates massive amounts of data. While this may seem a liability, "mining" this data is much like telemetry from space launches, provides an unparalleled opportunity to identify and mitigate potential accidents and catastrophic event, which can greatly advance the state of rail safety.

While the level of data analysis does not approach the scope or complexity of the much publicized "big data" (e.g. [4]) it does retain much of its basic elements of data capture, analysis, search, and information privacy.

Data Analytics falls into three general categories: descriptive, predictive, and prescriptive. Descriptive analytics, often called "data mining" or "business intelligence" establishes the reasons for past successes and failures by analyzing data from past events. Predictive analytics, often known as "forecasting" builds on descriptive analysis, by using a variety of statistical tools / techniques such as data modeling and machine learning to analyze historical and current data to make predictions about future events. Such methods establish the likelihood (i.e. probability) of events, including preconditions, from which risks and opportunities can be better understood and enabling better decisions making processes. Prescriptive analytics builds on predictive analysis to develop decision options by synthesizing data, predictive analysis, and rules of business. Predictive analytics not only anticipates what events may happen, but when and ideally why such they occur. Moreover many prescriptive data analyses use multiple data sources, internal (e.g. organizational) and external (e.g. social media).

Understanding Data

Most people understand data to be quantitative or number based. Much of Metrolink's PTC system data are quantitative values: success / failure, time stamps, locations, and speed. Such data can tell the analysis that an event has happened as well as when, where, and what the locomotive or cab car was doing (accelerating, braking, over speed, etc.). However such data cannot establish the "how" or "why" of those events. For that, qualitative data is needed, which was why communicating with dispatchers, locomotive engineers and other observers is necessary to develop a contextual "narrative" of events. As this paper will illustrate a using good mix of quantitative and qualitative data can be much more effective that quantitative analysis alone.

Revisiting High Reliability in Organizations

This author has explained how high reliability organizations (HROs) in the rail industry could be developed by proper PTC implementation, [1] as well as through psychological principles and organizational culture [2]. In short, HROs must adopt not just tools but the underlying principles and philosophies that enable them. HROs can be further enabled using safety focused data analytics.

The principles of HROs were conceived to manage tightly scheduled operations while maintaining low risks within inherently high-hazard environments, using numerous organizational processes. HROs instill inherent climates of safety to allow organizations to "... repeatedly accomplishes its high hazard mission while avoiding catastrophic events, despite significant hazards, dynamic tasks, time constraints, and complex technologies." [1]

Becoming an HRO is <u>not a goal, but a process</u> characterized by 5 basic principles [1]:

1. Preoccupation with failure
2. Reluctance to simplify interpretations
3. Sensitivity to operations
4. Commitment to resilience
5. Deference to expertise

Process activities that can help develop HRO are [1]:

1. Develop a system of process checks to spot expected and unexpected safety problems
2. Develop a reward system to incentivize proper individual and organizational behavior
3. Avoid degradation of current process or inferior process development
4. Develop a good sense of risk perception
5. Develop a good organizational command and control structure.

The aviation industry has used automation over the past 30 years to improve safety. PTC, as a form of automation, promises improved safety as well, but the rail industry must recognize that the aviation industry's successes have resulted by steadfastly incorporating HRO principles and processes as it continued to adopt automation, as well as through learning from heartbreaking cases like Air France Flight 447 and Asiana Flight 214 [3].

Brief Review of Psychological and Culture Factors in Organizations [2]

HROs require organizational cultures – like natural environments – conducive to desired outcomes, i.e. safety. Properly managing the underlying psychological and cultural factors that develop and sustain organizational environments, including intrinsic (internal) and extrinsic (external) factors that drive worker actions within a working environment, is essential to promoting robust, healthy organizations that operate efficiently and safely [2].

Underlying employee motivations are a key element of interactions. One extreme (Theory X) considers employees inherently lazy and needing supervision, while the other (Theory Y) theorizes they will be self – motivated. Employees are further motivated by whether or not basic (e.g. food or shelter) or more esoteric needs are met (e.g. equity or sense of purpose). Employees also exhibit certain personality temperaments like "by the book" or "analytical," and maintain levels of ethical standard from punish / reward to internalized standards [2].

Lastly, organizational culture (i.e. "mental software or collective programming) are normalized behavior within the organizational environment characterized by five major factors: Power Distance, Masculinity / Femininity, Individualism, Uncertainty Avoidance, and Long / Short Term Orientation [2].

Learning from Other Industries

There several good examples of using data analytics and lessons learned from which the rail industry can draw. One of the more notable is of course the NTSB's Accident Reports site [5]. While this site is not strictly a data analytics resource, it provides many of the lessons learned that are the heart of many data analytics effort, including developing rules for prescriptive analytics. The SKYbrary is a similar source that acts as a reference for aviation safety [11].

NASA has developed the Aviation Safety Reporting System (ASRS), which allows frontline personnel in the aviation industry to contribute anonymously to the world's largest repository / database of voluntary, confidential safety information [13]. The database includes narratives (with identifying details removed) that provide "an exceptionally rich source of information for policy development, human factors research, education, training, and more."

NASA's Public Lessons Learned System, a searchable database site that contains official, reviewed lessons learned from NASA programs and projects. These lessons describe the original driving event and offer recommendation that "feed into NASA's continual improvement via training, best practices, policies, and procedures." [6] The FAA's Lessons Learned from Transport Airplane Accidents site follows a similar format to NASA's, but focuses on the aviation industry [7].

Sites related to more traditional data analytics are the FAA's Aviation Safety Information Analysis and Sharing (ASIAS) System [8] and FAA's Data and Research Website [9]. Both sites offers advanced analytical capabilities including integrated queries across multiple databases and extensive warehouses of safety data.as well as narrative lessons learned data from which to draw rules for prescriptive analysis.

A more advanced ASIAS site includes millions of flight data records from both public and non-public (de–identified) aviation data sources. It plans to become a hub for data exchange and analytical capabilities among a network of at least 50 domestic and international airlines over the next few years [10]. The site is directed by both government and industry representatives Public data sources include, air traffic management data related to traffic, weather, and procedures. Non-public sources include de-identified data from air traffic controllers and aircraft operators, digital flight data and safety reports submitted by flight crews and maintenance personnel [10]. Analyses included directed studies, assessment of safety enhancements, known risk monitoring, and vulnerability discovery, the results of which analyses are shared with ASIAS participants. ASIAS also established key safety benchmarks to allow individual operators to assess their own safety performance against the industry as a whole [10].

One of the more interesting developments has been NASA's System-Wide Safety and Assurance Technology (SSAT) Project which used natural language reports by pilot and air traffic controller reports for trend to predict potential risks based on natural language reports. However this project has not continued investigations since about 2012.

Adapting Safety Reporting Systems to the Rail Industry

The Federal Rail Administration (FRA) and other participating railroad carriers and labor organizations partnered with NASA to create the Confidential Close Call Reporting System (C3RS) [14], which adapts the qualities that have made the ASRS successful. The C3RS program includes [15]:

1. Voluntary confidential reporting of close-call events by employees;
2. Root cause-analysis problem solving by a Peer Review Team composed of labor, management, and FRA;
3. Identification and implementation of corrective actions;
4. Tracking the results of change; and
5. Reporting the results of change to employees.

The great innovation that C3RS brings to the rail industry is providing confidential reporting, and enabling labor, management, and FRA to jointly do "root–cause problem solving" [15]. Early C3RS evaluations indicate its great promise at increasing safety [15, 16].

The FRA has also partnered with multiple rail collaborators, to create Clear Signal for Action (CSA). The CSA Program includes [15]:

1. Voluntary peer-to-peer feedback in the work environment on both safe and risky behaviors and conditions (data is owned by labor and not disclosed to management);
2. Labor Steering Committee root cause analysis and the development of behavior and condition related corrective actions;
3. Steering Committee implementation of behavior related corrective actions;
4. Joint labor management Barrier Removal Team refining condition-related corrective actions and implementation;
5. Tracking the results of the change; and
6. Reporting the results of change to employees.

CSA's provides to the rail industry the innovation of peer–to–peer feedback on safe and risky behaviors and conditions, root cause analysis, and cooperation between labor and management in corrective actions [15]. Results for the CSA have been even more promising than the C3RS program, as the FRA considers the CSA program ready for industry– wide implementation after completing three pilot demonstrations.

The C3RS and CSA both enable many of the concepts and principles of a Risk Reduction Program (RRP) advocated by the FRA:

1. Proactive identification of hazards and risks;
2. Analysis of those hazards and risks;
3. And implementation of appropriate action to eliminate or mitigate the hazards and risks

which are key elements of HROs. However, while the FRA believes implementing C3RS or CSA would help a railroad develop an RRP, the FRA will not require any railroad to implement either [15]. Instead the FRA will promote such programs within industry as part of an organizational RRPs. As we will see, although Metrolink is not explicitly implementing C3RS / CSA, their efforts meet, if not exceed much of the FRA's vision for C3RS / CSA type RRPs.

NOMENCLATURE

ASIAS – Aviation Safety Information Analysis and Sharing
ASRS – Aviation Safety Reporting System
BOS – Back Office Server
C3RS – Confidential Close Call Reporting System
CSA – Clear Signal for Action
FAA – Federal Aviation Administration
FRA – Federal Railroad Administration
HRO – High Reliability Organization
NASA – National Aeronautics and Space Administration
PTC – Positive Train Control

RRP – Risk Reduction Program (RRP)
SSAT – System-Wide Safety and Assurance Technology
TMC – Train Management Computer
TMDS – Train Management Dispatching System
WMATA – Washington Metropolitan Area Transit Authority.

CASE STUDY EXAMPLE

Metrolink is one of the few FRA–regulated railroads that has successfully met the 2015 PTC implementation deadline set by the Rail Safety Improvement Act of 2008. While their herculean efforts to meet the deadline are highly commendable, one of the most promising developments of their implementation has been their innovative use of the data their PTC system generates, including near real–time analysis. Even more amazing is that the team consists of four Metrolink employees, only two of whom can be considered dedicated data analysts, in cooperation with dispatchers and engineers reporting their activities running Metrolink's PTC enabled fleet. This is highly encouraging as they have not yet had to employ a dedicated team of data scientists nor the sophisticated data processing algorithms, commonly associated with big data or data analytics, but have shown impressive preliminary results.

Basic Systems View of Data Generation

A back office server polls for data – "message" packets – which is returned by an onboard train management computer (TMC) placed on each train. Each message includes a train's location and speed, as well as a time stamp. Messages are transmitted also to the dispatcher / dispatch office and communication personnel, and stored by an onboard event recorder. The back office server (BOS) monitors onboard units, representing an individual train, and wayside components to create a real time map of its jurisdiction so as to coordinate movement throughout.

Metrolink reported that "regular" operations as long as the maximum lad time (t) between messages is less than 3 seconds, operations are recorded as "regular." In the event the TMC fails to return a message within 3 seconds, the BOS records a penalty and the time and location of the "cutoff." The BOS also reports the cutoff to the dispatcher. The PTC system would respond to the cutoff by issuing a stop command.

One of two reports is made for the time a train operates using PTC. *Success* indicates no incidents occurred while operating, i.e. the train departed active and finished active. *Failure* indicates an incident (e.g. "cutout") that is logged by the BOS. A computer generates a ticket indicating a communication problem and sends the log to the entire development team even during an intended test event. The TMC reports the incident to the engineer who informs the dispatcher to tell the back office to enter details into the PTC log about what happened in the field. Trains can also operate as a "missed opportunity" either by not running, initiating, or going active during initial testing of PTC.

Diagnosis and Recovery

Once incidents are logged, the back office staff begins a multi-level root cause analysis. They first examine the data to establish the nature of the system "failure." At present, Metrolink uses spreadsheets to record PTC data annotated with notes from engineers and other sources. Their simple but comprehensive solution allows them to categorize incidents, particularly cutouts, as typical, unusual, or not common, which enables them to establish a recovery strategy based on whether such events have been previously encountered.

The back office staff can check onboard and wayside communication problems remotely and restart devices as needed. In cases when the problems cannot be detected remotely and it is safe to proceed, the engineer, with dispatch approval, can manually cut out the on board system and proceed at reduced speed while in the block where the failure occurred.

When hardware problems cannot be solved remotely, a System Anomaly Report is created linking the problem to the hardware unit in the log until it is fixed or replaced. All incident reports are archived on the server daily at 12 midnight.

Current Performance

Metrolink currently operates about 165 trains per day. Current response time for these trains, from initial report of a PTC related problem to resolution, averages about 1 hour. This includes about 30 unusual PTC related occurrences per day. This translates to a decent operating capacity of 80 – 90% and a high of about 90 – 92% during initial the testing phase from January 2015.

Culture Change

Dr. James Reason has noted five essential components to organizational cultures that enable safety: ***informed, reporting, learning, just, and flexible*** [12]. Metrolink's efforts to "tame" PTC data is an amazing narrative of their efforts to enable safety above and beyond technological innovation.

Typical to most organizations with long traditions, there were levels of oppositions between engineers, trainers, and engineers developing and testing the PTC hardware and software. One of the most pronounced conflicts was a conflict between locomotive engineers and trainers, who each strongly maintained that the other was to blame for system failures, including cutouts. Further investigation noted that locomotive engineers failed to understand that PTC generated comprehensive data about the incidents that would often conflict with their accounts.

Engineer actions are not altogether surprising, as this author noted the role of psychological and cultural effects in organizational operations in [2]. In particular, the rail culture has often been criticized as being overly punitive, greatly affecting Reason's "**Just Culture**" component of safe cultures. It can further be argued that this earlier culture influenced engineers to perceive PTC infractions as threats to be avoided at all cost. Fear of reprisal is also one of the principal reasons why the ASRS and C3RS stress confidentiality, and the CSA stresses peer–to–peer review rather than direct interaction with management.

The conflict started to resolve once engineers began to understand how the PTC system worked and realized they were expected to report the details of why the system reported a failure, rather than being punished for it. At this point, communications with engineers greatly improved as engineers began to report details much more readily about PTC incidents, which signals a transition towards a better "**Reporting Culture**" and enables an "**Informed Culture**" created by the interaction of engineers and PTC system developers / maintainers.

Metrolink's "**Learning Culture**" has been greatly enhanced using aforementioned spreadsheets that record PTC data annotated with notes from engineers and other sources. This allow them to capture previous incidents and as we will see to identify and contend with potential catastrophic events.

Organic RRP Development

One of the most difficult elements of system safety is effectively illustrating that an accident was prevented by effective safety cultures. However, positive C3RS and CSA test results and the success of similar systems in aviation like the ASRS, indicate that reporting systems can reduce accident by increasing overall safety.

An interesting aspect of the cultural change about which this author wrote, is that the subsequent increase in trust and process transparency between locomotive engineers and computer engineers appear to have organically developed most of the elements found in C3RS and CSA. For example, while Metrolink's PTC system was seen initially as an automated informant, it has evolved into an identification tool that signals the need for root–cause analysis problem solving by a peer review team consisting of front line personnel, particularly locomotive engineers and developers. In addition the need for a labor – management barrier removal team are made redundant in the new culture as engineers are no longer passively investigated, but take active roles in the safety development process. The close collaboration effectively eliminates the need for reporting results as those most directly affected are part of the identification, analysis, and change process. Finally developers have developed a change tracking / lessons learned system using spreadsheets. As a result, Metrolink has effectively established a RRP process that proactive identifies hazards and risks, analyzes them, and, implements appropriate action(s) to eliminate or mitigate the hazards and risks through its proactive process.

An Example of Preemptive Accident Avoidance

The potential of Metrolink's Proactive Data Analysis is illustrated by a recent case study. In mid–2015, the PTC system commanded a train to stop for no discernable reason. Careful analysis revealed that the stop resulted from the engineer entering the wrong track when initializing the TMC. The Global Positioning System has an uncertainty of about 16 feet, about equal to that of many gauge and tracks running side by side in the field. In this case, the uncertainty in GPS allowed the engineer to select the wrong track ID on the TMC and operate on a different track altogether. The system had stopped the train because while there was a green signal on the block where the train was actually operating, the adjacent track block had a red signal. It was quickly recognized that PTC had operated correctly given the information it had and the limitations of GPS. However had the signals been reversed, where a red signal been on the correct track block and a green signal been on the adjacent track, conditions like that of the September 12, 2008 Chatsworth accident could exist with great potential for a similar crash. Similar errors where engineers select not only wrong tracks, by wrong weights which could cause PTC to miscalculate stopping distance were also identified through analysis of PTC incident logs.

To resolve the issues, engineers and developers cooperated to find a human factors solution which was formulated initially by an engineer known as DK. DK observed that when engineers select the track, the Train Management Dispatching System (TMDS) transmits a message to the BOS as well as the dispatcher. DK suggested the procedure be changed to require engineers to match the track selection message data with what they knew to be the actual track they were on. If the engineer cannot confirm a match, the engineer asks the dispatcher to confirm whether they are on the correct track. If not, the system assumes the input is incorrect. In all cases, the procedure assumes the engineer will select incorrectly and requires the engineer to establish they have selected correctly.

The Future of Proactive Data Analysis in the Rail Industry

While Metrolink continues its efforts to utilize the additional data generated by PTC, it is apparent that great benefits can be achieved with little additional expenditure. This does not preclude the use of big data methods as more information becomes available for use in the railroad industry, but it should encourage rail industry decision makers to consider how data analytics can be used to improve safety, efficiency, on–time performance, and customer satisfaction, etc. concurrently.

It may be argued that Metrolink's efforts imitate the C3RS and CSA program, but they differ significantly. For one the process augments human observations using automation, because reports are generated by the PTC system in addition to supporting data.

Metrolink's organizational culture also significantly changed to adapt to the new processes needed to perform proactive analysis. The automation augmented process signals a need for system troubleshooting which may have been missed by human agents. A large number of adverse events result from lack of associating key factors before the fact. For example the 2009 Washington Metropolitan Area Transit Authority (WMATA), Fort Totten Crash in Washington, DC was preceded by issues with a faulty track circuit for nearly 18 months. However Metrolink's process actively treats troubleshooting as part of the larger safety process rather than separately. For example, the issue of entering wrong track information was identified after PTC generated a ticket. This triggered engineers and developers to identify, assess, and resolve the issue quickly, and illustrates how RRPs and HRO appear to evolve naturally from Metrolink's innovation.

As problems are resolved and logged, the PTC database / spreadsheet grows iteratively, therefore recurring issues are identified whenever a ticket is generated and found to already have been associated with earlier repairs. Such lessons learned could be crosslinked to safety related issues to develop more comprehensive and informed lessons learned. It can also provide a wealth of information should steering committees that were a part of CSA be implemented, particularly at a level that affects the entire rail industry

Metrolink's innovative efforts show great promise advancing the safety of the railroad industry using relatively simply technology in cooperation with engineers who add context to data. This author therefore encourages the FRA to study further the potential and efficacy of Metrolink's efforts and how other rail companies implementing PTC can learn from Metrolink's experiences utilizing the data from their PTC systems to examine PTC system health. This author's also thinks that further research be conducted applying other data analytic methods to the rail industry as a whole. Such systems used in conjunction with human reporting systems like C3RS and CSA have great potential to establish safety in depth within the rail industry.

REFERENCES

[1] Placencia, G, Meshkati N, Moore, J, Khashe, Y. (2014). Technology and High Reliability Organizations in Railroad Operations Safety: A Case Study of Metrolink and Positive Train Control (PTC) Implementation. *Proceedings of the Joint Rail Conference*, 2014, Colorado Spring, CO.

[2] Placencia, G. (2015). Psychological and Cultural Components Affecting Rail Worker Culture: A Literature Review. *Proceedings of the Joint Rail Conference,* March 23–26, 2015, San Jose, CA.

[3] Placencia, G. (2015). Positive Train Control: Developing Lessons Learned About Automation and High Reliability from the Aviation Industry. *Proceedings of the AREMA Conference*, October, 2014, Minneapolis, MN.

[4] Zarembski, A. M. (2014). Some Examples of Big Data in Railroad Engineering. 2014 IEEE International Conference on Big Data. Oct. 27–30, 2014, Washington DC. pp. 96 – 102. DOI: 10.1109/BigData.2014.7004437

[5] NTSB Accident Reports Website. http://www.ntsb.gov/investigations/AccidentReports/Pages/AccidentReports.aspx. Accessed November 30, 2015

[6] NASA Public Lessons Learned System (LLS) http://llis.nasa.gov/. Accessed November 30, 2015.

[7] Federal Aviation Administration: Lessons Learned From Transport Airplane Accidents. http://lessonslearned.faa.gov/. Accessed November 30, 2015.

[8] Federal Aviation Administration: Aviation Safety Information Analysis and Sharing (ASIAS) System. http://www.asias.faa.gov/pls/apex/f?p=100:1: Accessed November 30, 2015.

[9] Federal Aviation Administration: Data and Research Website. http://www.faa.gov/data_research/. Accessed November 30, 2015.

[10] ASIAS Industry Collaboration Website. https://portal.asias.aero/web/guest/home. Accessed November 30, 2015.

[11] SKYbrary Website. http://www.skybrary.aero/index.php/Aviation_Safety_Statistics. Accessed November 30, 2015.

[12] Reason, J. (1998) Achieving a safe culture: theory and practice. *Work & Stress*, 1998, Vol. 12, No. 3 293-306.

[13] Aviation Safety Reporting System Website. http://asrs.arc.nasa.gov/. Accessed January 22, 2016.

[14] NASA Confidential Close Call Reporting System Website. http://c3rs.arc.nasa.gov/. Website. Accessed January 22, 2016.

[15] Federal Register (2015). Vol. 80, No. 39, February 27, 2015.

[16] Clear Signal for Action Program Addresses Locomotive Cab Safety Related to Constraining Signals. (2007). U.S. Department of Transportation, Federal Railroad Administration, RR07-08, February 2007.