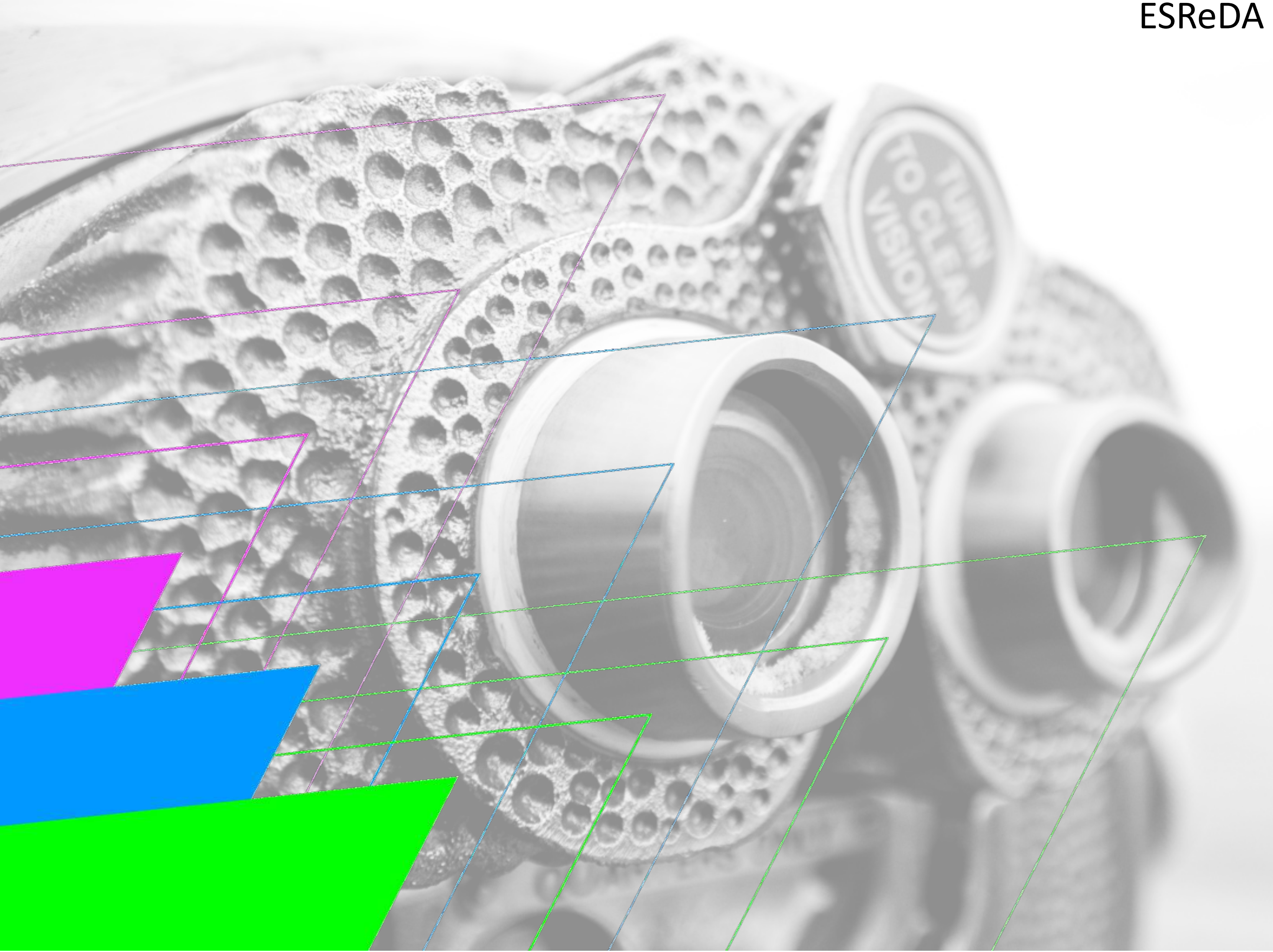
Enhancing Safety: The Challenge of Foresight

ESReDA Project Group *Foresight in Safety*



Chapter 10

Big data analytics and early warning signs

Eric Marsden
Nicolas Dechy
Ana Lisa Vetere Arellano

ESReDA
European Safety, Reliability & Data Association

Table of contents

10.1 **Executive summary** 181

10.2 **Key messages** 181

10.3 **Context** 181

10.4 **Motives and benefits of big data analytics** 183

10.5 **Safety and security applications of predictive analytics** 183

10.5.1 Detecting new safety and security threats 184

10.5.2 Monitoring effectiveness of safety barriers 185

10.5.3 Safety investigation 185

10.5.4 Condition-based maintenance 186

10.5.5 Structural health monitoring 187

10.5.6 Fraud detection 187

10.6 **Challenges and risks to the effective use of big data analytics** 187

10.6.1 Level of confidence in predictions 187

10.6.2 Data silos and the challenge of data interoperability 188

10.6.3 Lack of expertise 189

10.6.4 Black boxes 189

10.6.5 Man-machine interface 190

10.6.6 Incomplete data – formal versus informal, tacit, and quality 190

10.6.7 Data analysis 191

10.6.8 Machine learning biases 191

10.6.9 Data governance and ethics 191

10.6.10 Invalid conclusions 192

10.7 **Conclusions** 192

10.8 **References** 193

10 Big data analytics and early warning signs

Eric Marsden, Fondation pour une culture de sécurité industrielle (FonCSI), France,
Nicolas Dechy, Institut de radioprotection et de sûreté nucléaire (IRSN), France,
Ana Lisa Vetere Arellano, European Commission Joint Research Centre, Ispra, Italy.

10.1 Executive summary

The analysis of "big data" has generated a large amount of interest in industry over the past decade. Promoters can point to a number of benefits and success stories in the retail and entertainment areas, as well as in optimization of industrial processes, and more recently in overcoming security (Amanullah et al., 2015), safety (Huang et al., 2019) and reliability (Ham & Park, 2020) challenges.

To what extent can increased collection and analysis of data help to detect early warning signs of system failure, and predict the occurrence of the infrequent events that are relevant to the management of industrial safety and major accident hazards? What should we monitor? What are the challenges and risks of inadequate use of big data?

10.2 Key messages

Big data analytics has a significant potential to improve the detection of early warning signs of system failure, when compared with the use of standard statistical tools or of unassisted human monitoring of system data. It also provides opportunities in health usage monitoring of equipment (e-Health).

Industry 4.0 has fostered the burgeoning of digital twins, offering enterprises the capability of real-time situation awareness thanks to digital technologies (ubiquitous sensors generating big data, artificial intelligence, machine learning, predictive software analytics, etc.). These dynamic digital simulation models of physical assets allow continuous data, information and knowledge (DIK) acquisition related to system performance and failure. Thus, enterprises are able to rapidly run diagnostics, intervene to correct problems before they become critical and detect new improvement opportunities.

Effective use of predictive analytics faces a number of obstacles, some induced by the black-box-effect of algorithms and tools. Their use requires specialised skills of the analysts who build the data collection and analysis tools, but also from the users, who will require training though they will not become data analysts.

Algorithms can improve human judgment, but will never replace it completely. Humans, within organisational processes, will remain important at every step of a big data analytics process: framing individual and organisational attention to the data that important to analyse, putting in place and checking the data collection process, interpreting the importance of outliers found in data, and validating the causal nature of correlations identified.

Automatic decision-making based on algorithms should be very carefully controlled to check for potential biases in available data and its treatment.

The big data paradigm emphasizes the importance of data *quantity*, but analysts should start by checking the *quality* of data and its relevance to the analysis undertaken. This issue is particularly significant when addressing the human and organizational dimensions of system operation and searching for the underlying causes of events. Moreover, the big data paradigm focuses on official reported data, disregarding tacit and informal data, which is sometimes critical to understanding the complexity of social and organisational issues that affect safety.

Data and information are constructed by worldviews, tools such as sensors and human perception; different meanings can be associated to the same data or word by different people and groups of professionals.

Big data analytics raises questions regarding privacy, information security, ethics, which should be handled by a well-designed data governance process.

10.3 Context

Statistical analysis of reliability data has a long and illustrious history, including Dr Snow's analysis of the 1854 cholera epidemic in London, which identified the public water fountain at the epicentre of the outbreak. Research over the past 50 years has also shown that even simple mathematical models, such as linear regression models, provide better predictions and forecasts than human experts in a range of situations[39], by avoiding a range of cognitive biases that affect human information processing.

Over the past decade, a new strand of work on "big data analytics", which applies statistical techniques and more recent machine learning techniques to larger sets of data, has gained much attention. The emergence of this concept has been enabled by a number of coinciding factors:

- Affordable computer systems that are able to store and process very large volumes of data.
- Increased use of "smart sensor" systems that can provide real-time data on various performance measures of equipment (temperatures, pressures, vibration levels, etc.). This "internet of things" trend has been enabled by technological advances in integration levels for microelectronics, by reduced power consumption and improved battery technology and by the development of low-power wireless communication technologies. Other sources of data include social media platforms, e-commerce and smartphones with geolocalisation features.
- Increased industrial use of sophisticated machine learning techniques[40] such as neural networks that allow classification, anomaly detection, and optimization.
- Development of natural language processing (NLP) tools that allow automated treatment of large volumes of unstructured text. These tools are often based on statistical learning rather than on language models developed by humans. Companies often possess large corpuses of unstructured text, including historical information which has not been manually classified when companies moved from paper-based to computer-based storage of reports and data logs. Until recently, unstructured text could not be used in analytical tools without significant manual effort to classify the data. New techniques can extract structured information from these data sources and allow their combination with numerical data from other sources. NLP also enables automatic classification or clustering of documents according to their level of similarity.

Table 1: Comparison between structured and unstructured (with natural language processing) data analysis (Dechy and Blatter, 2018)

	Prior assumptions to verify	**No prior assumptions to verify**
Structured data analysis	Validation of the correlations identified by experts	Cross-referencing databases to identify targets to be analysed Cluster analysis to discriminate
Analysis of textual data extracted from event reports	Automatic semantic categorisation of incidents	Failure analysis to identify the features of the systems considered: type, manufacturer, composition,…

The term "big data" is generally used to refer to new generations of data collection and analysis systems, which were initially characterized by "three Vs":

- a large **volume** of data[41], that typically cannot be stored on a single personal computer;
- high **velocity**: data sources that generate large streams of events that cannot feasibly be stored, but must be filtered and analysed in real time;
- significant **variety**: different data formats, often unstructured or multimedia, which are difficult to store in traditional relational databases.

Four other V's have later been added to this motto (Khan et al., 2014):

- **veracity:** large volume and flows of data are automatically collected, but they may be erroneous or their accuracy becomes harder to check;

[39]The field known as "clinical vs. statistical prediction" was developed by psychologist Paul Meehl, who reported on 20 studies that compared the predictions of well-informed human experts with those of simple predictive algorithms. The studies ranged from predicting how well a schizophrenic patient would respond to electroshock therapy to how likely a student was to succeed at college. Meehl's study found that in each of the 20 cases, human experts were outperformed by simple algorithms based on observed data such as past test scores and records of past treatment. Subsequent research has decisively confirmed Meehl's findings: more than 200 studies have compared expert and algorithmic prediction, with statistical algorithms nearly always outperforming unaided human judgment.

[40]Machine learning is based on algorithms that can "learn" (infer relationships) from data without relying on rules-based programming.

[41]As an illustration, a typical offshore oil rig can [generate two TB of data per day](), and an A350 aircraft includes more than 6000 sensors (temperature, pressure, operating speed, stress, humidity, vibration, fuel flow rate, etc.) that generate more than 2.5TB of data per day.

- **variability:** the rate of change of the structure of the data, which depends on the stability of the context in which data is extracted (for example, the same word used by the same person in different contexts and tones of voice may signify different meanings);
- **value:** the net added value for users, which is the difference between the gross benefit and the cost of collection and analysis;
- **visualisation**: the quality and relevance of data visualisation is a key to reveal the significance of data analysis and control biases of representation.

10.4 Motives and benefits of big data analytics

Some analysts, start-ups and industrial promotors, full of optimism concerning the potential of these technologies and techniques to improve industrial production, refer to a fourth industrial revolution, or "Industry 4.0", in which continuous streams of real-time data from sensors within the production line and upstream supply chain can be analysed using artificial intelligence techniques to allow increased product customization, performance optimisation, more flexible and adaptive production and anticipatory detection of critical events.

Leveraging these sources of data to improve decision-making (an activity called *predictive analytics*) requires a combination of skills in new computer technologies, statistical analysis, machine learning and data visualisation (an intersection called *data science*). Their application for reliability and safety purposes is more recent (10 years ago for software from start-ups).

Traditional collection and analysis of safety data (operational experience feedback, event reporting, generation of safety performance indicators) have long been used for safety management, to identify anomalies and to check that interventions result in a system improvement. Reliability engineers have long used trend analysis on critical system measures such as operating temperatures or pressures to identify deviations from design levels and from normal operating conditions. When thresholds are passed, alarms inform plant operators; when extreme levels are passed an emergency shutdown is triggered. Big data analytics, when applied to safety issues, is a complement to these traditional methods that analyses more complex correlations or interactions between multiple variables to help identify more subtle anomalies, that can have performance or safety implications. These more sophisticated analysis techniques are also better able to account for slow changes in plant performance as it ages than the static thresholds used in traditional trend analysis. They may be able to produce relevant insights when the plant is operating outside standard conditions (for example during startups and shutdowns), which is rarer for traditional statistical analysis methods.

Traditional statistical analysis of data requires the analyst to formulate a hypothesis, then collect relevant data and undertake an analysis to check whether the data supports the hypothesis. The specific promise of new "unsupervised learning" models is that a computer might identify 'patterns', 'features' or 'a model in the data' that predict specific outcomes (such as mechanical breakdowns or technical failures) in an "automatic" manner, without benefiting (or indeed suffering[42]) from the preconceived assumptions of the safety analyst.

As discussed in section 10.6, this promise does not eliminate the need for an experienced analyst to examine the features identified by the algorithms, verify the assumptions and assess whether they are of relevance to operational performance or safety in general and in a specific context of use[43].

10.5 Safety and security applications of predictive analytics

The most prominent applications of predictive analytics techniques have been in e-commerce and entertainment, with recommendation engines ("if you liked that, you might like this") and targeted advertising ("if you searched for this, you might buy this") being the most commonly developed features. The techniques also have applications in the safety domain, related to the detection and analysis of early warning signals. In particular, the algorithms, processing facilities and approaches can be applied during:

[42] *Big data analysis sometimes generates insights that contradict experts' prior knowledge. For example, a large mining company undertook a clustering analysis to identify which of 620 data points and metrics concerning employees was correlated with workplace injuries and fatalities. Some of the findings, such as the fact that most incidents occurred less than half a day into a shift, or that highly tenured employees had significantly higher accident rates, challenged managers' views of the drivers of accidents.*

[43] *The "knowledge hierarchy" analysed in the knowledge management literature distinguishes between data, information, knowledge and wisdom (Rowley, 2007), and provides a framework to describe the processes involved in moving from a low-level element in this hierarchy, such as data, to a higher-level element, such as information.*

- *The detection phase of an experience feedback/reliability analysis process:* they can help to detect anomalies, unusual trends, typical configurations and emerging patterns of behaviour that may affect only a subset of equipment or a population of system users. This can be used to detect the early warning signs mentioned in chapter 6 (Strucic, 2020) dedicated to the visibility of early warning signs.
- *The analysis phase:* they can help analysts to dig deeper and test their hypotheses. The "slice and dice" data processing facilities that are associated with the development of a big data infrastructure can be used by analysts to extract all events that match specific search criteria, to filter issues and to check for trends or anomalies in these events, in a much more convenient manner than when data was fragmented across multiple departments and storage systems.
- *The prediction phase:* they can help safety experts anticipate the performance of system changes before they are implemented, through more sophisticated and precise system models.

Data sources that can be used for these safety applications of predictive analytics techniques include:

- Data generated by equipment and sensors, with very high volumes and high frequency of use in modern technological systems where many components have been instrumented to generate monitorable outputs.
- Operational data from management systems, such as the number of inspections undertaken, number of flights flown, number of customers.
- Text written by humans, such as the descriptions included in incident reports and inspection reports, the content of emails.

In the security domain, predictive analytics is being used in the following ways:

- *Maintenance of military vehicles*: predictive maintenance allows the military to reduce malfunction and failure of vehicles in operation, thanks to real-time data collection and analysis from sensors and telematics[44].
- *Prediction of soldier effectiveness*: in a virtual near-reality environment, soldiers can be monitored to predict how they will react with the help of biosensors that collect real-time data that can be analysed by predictive analytics and machine learning algorithms[6].
- *Tracking readiness of equipment*: in order to better manage military training and defence operations, real-time access to intelligence related to the degradation state and location of equipment is essential in order to make better informed decisions, e.g. if a tank needs to be moved from one military base to another[45].

Big data analytics techniques allow a move from a static analysis of the factors of system performance to a dynamic and continuous approach, allowing more customisation to the specific characteristics of the system.

In the next paragraphs, we describe a number of applications of predictive analytics to safety management.

10.5.1 Detecting new safety and security threats

Collecting and treating massive quantities of data may allow the early detection of anomalous situations which may represent new component failure modes or threats to system safety. Big data analysis techniques allow high-dimensional data to be analysed using data mining techniques, searching continuously for new correlations or new outliers between multiple streams of data, such as those provided by various sensors (temperature, pressure, flow rate, displacement, rotational speed, stress, vibration, concentrations, geographic or spatiotemporal location, etc.). This work can help analysts to identify and define new early warning signs, surprises and potentially problematic assumptions. After investigation to assess their relevance, these new features can be added to the system monitoring framework.

These new technological promises are particularly relevant concerning technical data, but are also to some extent applicable to data concerning human and organisational factors of safety, such as data extracted from reports written in natural language. Correlations and outliers identified also require a cautious investigation approach.

[44] *https://emerj.com/ai-sector-overviews/predictive-analytics-in-the-military-current-applications/*

[45] *https://defensesystems.com/articles/2018/06/19/comment-dod-analytics.aspx*

An example application of this approach in information system security is given by intrusion-detection systems, which monitor network data and machine usage patterns for signs of new security (malicious use) threats.

Big data for detecting hidden correlations and Early Warning Signs – lessons from an IMdR project

The literature review undertaken for this project[46] suggests that weak signals are not intrinsically "weak" (Guillaume, 2011; Jouniaux et al, 2014). Rather, the notion of weak signal is an extrinsic property of an observation; it requires links between the observation and other information sources to be established, like a pattern in a puzzle. The interpretation process needed to establish a weak signal involves 3 steps: (1) *detection*: identification of a link between one observation and a scenario that impacts risk, (2) *relevance*: qualification of the link between the scenario of impact to risk and the risk modelling, (3) *amplification*: confrontation of the weak signal with safety objectives and means to deal with it. Some strong signals can be weakened as they lead to no changes or actions. A number of accident scenarios (the Concorde crash in 2000, Three Mile Island in 1979, Air Moorea crash in 2006, Paddington rail crash in 1999) were revisited with these principles. A big data case study was tested on a database of several tens of thousands of incidents with sixty fields of data to describe incidents. Pre-treatment of data using principal component analysis enabled to reduce to five the number of relevant parameters to search for correlations. The algorithm based on random forests (a classification algorithm based on decision trees) was able to confirm dominant parameters but also identified a number of correlations that surprised experts. The experts were unable to understand the underlying causality relationship or why some specific system state emerged, but the big data treatment provided a new line of investigation or assumption to be verified.

[46] *Project P12-1 (2013) - Institut pour la Maîtrise des Risques, a French NGO, www.imdr.eu*

10.5.2 Monitoring effectiveness of safety barriers

Big data allows organisations to measure and monitor the effectiveness of individual barriers, using data on operations (for example sensor data for physical barriers, and semi-structured text data such as incident reports for organisational barriers). This type of integrity analysis is not fundamentally different in nature from earlier work by reliability engineers and safety managers, though the use of larger quantities of data and more sophisticated statistical analysis techniques can improve the effectiveness of the monitoring.

Monitoring unsafe behaviours in Chinese underground mines

In order to effectively predict and decrease the number of "unsafe behaviours[47]" in Chinese underground coal mines, safety specialists analysed unsafe behaviour data of 2220 coal miners between 2013-2015 (Qiao et al., 2018).

Thanks to data mining techniques (association-rule and decision tree), the analysis of unsafe behaviours in underground coal mines helped to identify which unsafe behaviours needed to be better addressed to decrease frequency of accidents occurring. The study concluded that the factors that influence the frequency of unsafe behaviours were training (less training, more frequent unsafe behaviour), attendance (less attendance, more frequent unsafe behaviour), experience (less experience, more frequent unsafe behaviour) and age (very young and more elderly, more frequent unsafe behaviour).

10.5.3 Safety investigation

The data collected by equipment, if it is stored, can help safety investigators to understand the sequence of events that preceded the accident and to determine whether similar conditions have occurred in the past.

[47]. *This notion of "unsafe behaviour" is related to behavioural approaches of safety which have been criticized by workplace analysis of working conditions and real activities challenges within organisational constraints*

British Airways flight 38

A Boeing 777 crash-landed at Heathrow airport in 2008 due to a loss of thrust from both Rolls-Royce engines upon landing. The flight had taken a polar route, which led to the formation of ice crystals in the fuel. Upon landing, the temperature increase led to a slush of crystals flooding the fuel-oil heat exchanger and restricting the flow of fuel. The initial phases of the investigation found it difficult to identify the cause of the loss of engine thrust. "Data mining" techniques[48] were used to attempt to identify whether the flight had any specific features that differed from 175000 other Boeing 777 flights and which might explain the problem. The flight was found to be unique in combining low fuel flow during cruising and high fuel flow with low temperatures during approach[49]. A fix to the fuel-oil heat exchanger was implemented on all aircraft using these engines.

10.5.4 Condition-based maintenance

Traditional maintenance plans are either corrective (equipment is replaced once it fails) or time-based (maintenance is planned according to predefined component lifetimes based on statistical treatment of successes and failures of several component). The data collected from machinery and smart sensors embedded in a plant allows the implementation of another category of maintenance plan, condition-based maintenance, where replacements are planned depending on the degree of wear or corrosion of the specific pieces of equipment. Use of the approach improves the predictive ability of maintenance workers and allows them to optimize the availability of the component and the logistics of spare parts management.

A few examples illustrate applications of this approach:

- Aircraft engine manufacturers now collect large amounts of data from multiple sensors embedded in their engines[50], which is transmitted to ground-based engineering centres. The data allows them to detect problems requiring maintenance even before the aircraft lands (a process called "engine health management"). The complexity and importance of this data analysis leads to new business models where engine manufacturers retain ownership of engines and bill airlines per hour of engine operation, rather than selling engines outright.
- Power plant operators can be warned in advance of changes in the operating conditions of a unit that are not sufficient to trigger standard monitoring alarms (because they don't exceed predefined thresholds), but do indeed, through the presence of a correlation between unusually high and unusually low readings for example, point to serious problems that could have a safety impact.
- Railway operating companies receive real-time data from their rolling stock indicating the state of braking systems, batteries, compressors, doors, cooling equipment and toilets. The French national railway operator estimates that these tele-diagnostics technologies have allowed them to reduce maintenance costs by 20% and to improve availability.
- Predictive failure analysis in computer systems allows failure of system components to be anticipated by recording and analysing internal diagnostic indicators that are continuously produced by components such as storage drives, processors and fans. The Self-Monitoring, Analysis and Reporting Technology (SMART) mechanism available in most hard drives is an example of this process that is available even in consumer equipment.

EU Horizon 2020 SafeClouds.eu project

SafeClouds.eu is a project funded under the EU H2020 programme addressing "SOCIETAL CHALLENGES - Smart, Green And Integrated Transport". Participants were aviation stakeholders, airlines, IT infrastructure experts, universities, safety agencies and air navigation service providers. The project investigated the use of artificial intelligence (AI) techniques, such as deep learning and artificial neural networks, to analyse the precursors of safety events. According to the project coordinator, Paula Lopez-Catala, "*Understanding the precursors and potential risks that may lead to a safety incident is critical to complementing the traditional methods of monitoring safety, reviewing accidents and incidents and extracting lessons learned*". The techniques and algorithms were customised and tested to be effective in

[48] Data mining describes the exploratory process of finding patterns and knowledge within data. Predictive analytics then attempts to leverage that knowledge to make predictions about the future (attempting to forecast, anticipate, or infer).

[49] See UK Air Accidents Investigation Branch aircraft accident report 1/2020, available online.

[50] An aircraft engine in the Pratt & Whitney 1000G family (used in the A320Neo) includes more than 5000 sensors, generating data at a rate of 10GB/s.

every safety scenario identified, including unstable approaches to terrain warning, mid-air losses of separation, and runway safety.

10.5.5 Structural health monitoring

The integrity of mechanical structures can be monitored by collecting and analysing large numbers of measurements over time (temperature, pressure, vibration, strain, electrical conductivity, mass flow rates). Sophisticated condition monitoring and anomaly detection systems, monitoring real-time flows of data from smart sensors, can enable the safe life extension of aging industrial facilities. By analysing multiple sources of data, these systems can reduce the false alarm rate, which is a significant barrier to the implementation of simpler anomaly detection systems.

Lessons from an IMdR project on Health Usage Monitoring Systems

This project[51] identified four functions in implementing HUMS – Health Usage Monitoring Systems: (1) acquire and treat data from equipment, (2) diagnose the state of equipment by analysing flaws and failures observed, (3) establish a prognosis of the equipment state (4) aid decision-making based on current and foreseen evolution of state. Behind the technical vision of HUMS, the data format and software languages and interfaces, attention to organisational and human factors that influence the design of HUMS as well as their use, in operation, maintenance and logistics is a key. To calculate remaining useful life, some approaches rely on physical modelling of equipment (model-driven), some are data-driven, and some experience driven (based on expert judgment with inference to cognitive ontologies), while others combine the three. The prognosis should be established with regard to: its time horizon, the application domain, the level of decision, the perimeter, the freshness of data, the dynamics of the phenomenon, the level of detail and input data available. The project led to a practical guide that helps identify the questions to address with lessons from transportation (aviation and rail), the military and energy production sectors.

10.5.6 Fraud detection

Banks have long been using big data analytics to analyse large, unstructured data sets of transaction information and communication records to identify anomalous behaviour (internal fraud, credit card fraud, money laundering, etc.). Similar anomaly-detection techniques are used in the early stages of pharmaceutical research and drug development, with data mining techniques attempting to identify correlations between consumption of certain substances and health effects.

10.6 Challenges and risks to the effective use of big data analytics

In the following, we discuss a number of specific new challenges and recurring issues that safety analysts face in attempting to use predictive analytics to improve the detection and analysis of early warning signals. We also discuss ned risks generated by the use of big data techniques.

10.6.1 Level of confidence in predictions

Despite the high levels of interest and increasingly widespread implementation seen today, big data analytics are not a magical solution to all risk management problems. For example, a large quantity of information, both historical and real-time, is available on earthquakes, yet their prediction is extremely difficult (Silver 2012); the results of the 2016 and 2020 elections in the USA suggest that foresight concerning complex social systems is very difficult to achieve.

Failure in foresight on the Snorre Alpha oil rig

Over the past two decades, many companies in high-hazard environments have started to use safety climate surveys. These quantitative data analysis methods can be seen as an attempt to apply statistical methods to measure and analyse certain organisational factors of safety. As an illustration, staff operating the Snorre Alpha offshore oil extraction platform in the North Sea were subjected to a standard safety climate questionnaire in 2003, and analysis of the data indicated no points of concern. Only a few months later, the rig suffered a blow-out, an incident with potentially very severe

[51] Project reference P15-2 (2017) undertaken by the Institut pour la Maîtrise des Risques, a French NGO, www.imdr.eu

consequences. The resulting investigation identified a number of serious concerns with the way tradeoffs between production and safety were managed on the platform and whether the organisational culture encouraged staff to raise concerns (Antonsen 2009). This example suggests that the value of safety climate survey data in predicting safety performance is very low (worse, it may even encourage managers to develop a false confidence in their site's safety culture), though they may on occasion help understand certain organisational weaknesses[52].

10.6.2 Data silos and the challenge of data interoperability

For technical, historical, political and practical reasons, data is often generated and stored in "silos", or activity-specific information-processing systems which do not inter-operate. Companies operate a range of systems and equipment, each one specialized for a specific technical function and domain of expertise and for a different purpose (reliability, safety, purchasing); each produces data in a specific format and underlying data model. Establishing bridges between these data sources, i.e. making them interoperable, to enable cross-referencing and analysis of correlations, or moving to a unified "data lake" architecture (see Figure 1), is often a significant technical challenge to the effective implementation of big data programmes. It may also constitute a political challenge, because the ownership of data is a source of organisational power.

This challenge is compounded for companies that rely on numerous suppliers and contractors to design and assemble products, since relevant data is owned by large numbers of companies within the supply chain and network of partner organisations.

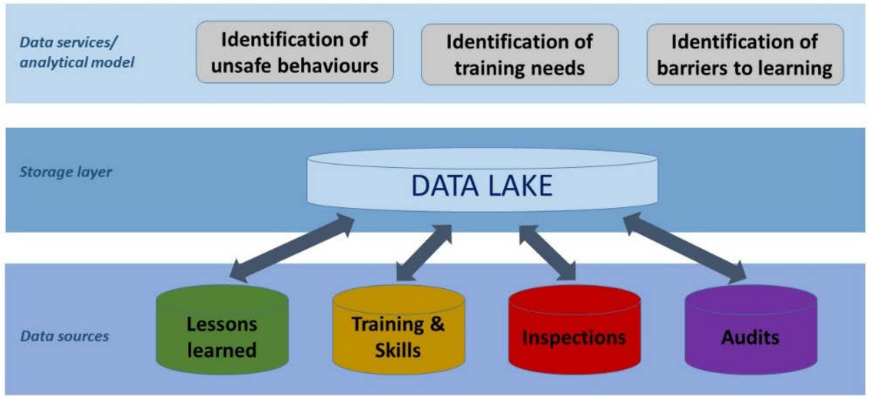


Figure 1. Example of data lake architecture.

Examples of interoperability challenges that companies could face when investing in a big data programme include (Scheerlinck et al., 2018):

- poor data quality;
- data protection considerations (confidentiality, privacy regulations);
- data sources with different data licences;
- requirement differences between data producer and user;
- difficulties in data integration, in particular when the number of data sources is high;
- rapid data integration difficulties linked to increased demand for near real-time analytics to generate insights in a timely manner;
- difficulties related to interfacing mechanisms between systems, e.g. incompatible communication protocols and data formats.

Further challenges include social dimensions such as vocabularies, different meanings and contextual backgrounds that co-exist with different sub-cultures (internally and externally).

[52] For instance, the Baker panel report, is an independent audit of the five BP refineries in the USA; it has published after the accident at the BP Texas City refinery the 23rd March 2005; it has used questionnaires to help understand, ex post, a number of organisational weaknesses affecting safety management on the five BP refineries in the United States. http://sunnyday.mit.edu/Baker-panel-report.pdf

EASA's Data4safety initiative for the European aviation sector

The Data4Safety partnership initiated by the European regulator, the European Union Aviation Safety Agency (EASA) in 2017 aims to collect and analyse data from many organisations involved in aviation safety in the EU, including airlines, airlines, air traffic management service providers, national aviation authorities, aircraft and engine manufacturers, weather agencies. The project will allow centralized analysis of data that is currently fragmented across a large number of organisations.

10.6.3 Lack of expertise

Effective use of data to improve organisational performance and safety requires a combination of skills in statistical analysis, machine learning, programming, the use of new storage technologies and data retrieval and query technologies. Analysts must also understand the system and its safety barriers to appreciate potential safety impacts; they must also possess communication and storytelling skills to present results in a form which is understandable by decision-makers. Although the situation is progressively improving, there still is a significant lack of specialists with such skills (Espinosa et al., 2019). Such skills are sought after by companies in many industries that are building up multi-domain and integrated data science teams.

Given this range of competencies, which are rarely held by a single individual, the collective process to design and operate a big data system requires cooperation between domains of expertise and territories of responsibilities.

To address this skill shortage, some companies are also resorting to training their own staff and changing their traditional *modus operandi*. Implementing big data and automated natural language processing techniques are not just a technological change with new tools for users. It will lead to change of practices and organisations across the different interacting disciplines (Rousseau et al, 2018). Some companies are reluctant to adopt a big data solution due to the significant investment required.

To build confidence in the predictions and algorithms, it will be necessary that domain experts and data scientists[53], work together on the data samples to develop the performance indicators, and make several tests.

10.6.4 Black boxes

Some of the new classes of algorithms used to analyse big data, such as neural networks, originate in the field of artificial intelligence. These algorithms implement a form of "machine learning", being trained on large quantities of input data to optimize some specific output quantity, such as the ability to recognize faces in images, to group observations into clusters, to identify anomalies in a stream of data. The resulting models are "black boxes", since —unlike classical statistical models — the analyst is not able to inspect the model to understand why one event is being classified in a specific way or why one point in a time series has been highlighted as anomalous. For example, if a neural network misclassifies an image, it is very difficult to determine which specific features in the image led to the neural network's error[54].

The opaque nature of these models has several drawbacks, when compared with simpler statistical models such as regression models:

- When a potential early warning signal is detected (for example by a neural network trained for anomaly detection), typical tools do not provide an explanation for why the situations is judged to be anomalous, providing little assistance to analysts who must attempt a diagnosis.
- They do not help analysts to build and test mental models of system operation, nor to make "rule of thumb" checks on model plausibility.
- Regulators have no way of checking the model's internal validity or underlying assumptions.
- The legal system cannot inspect the logic underlying the model's predictions, in case of an accident.

[53] *IMdR project n°17-4 (2019) concerned "big data and reliability".*

[54] *Consider for example a problem that affected a machine learning system developed to identify skin cancer in photographs of skin lesions. The system was trained on photographs of skin that were labelled by dermatologists, some malignant (affected by cancer) and some benign. Unfortunately, dermatologists often include a ruler in a photograph of a skin, to provide a reference of scale, and rulers were often present in the photos labelled "malignant" and absent in the photos labelled "benign". The machine learning system therefore learned to detect rulers [Esteva et al 2017]. See also https://www.technologyreview.com/s/601860/if-a-driverless-car-goes-bad-we-may-never-know-why/ for a description of why this may lead to problems understanding the behaviour of autopilots in vehicles.*

- Model predictions (related to the effects of implementing a new safety barrier, for example) may obtain less buy-in from decision-makers, because they do not help identify a plausible cause-effect mechanism and develop intuition concerning system operations.

Research into “explainable AI” (Hagras 2018) attempts to resolve these challenges related to transparency, lack of bias and fairness in the application of artificial intelligence techniques to decision-making in the public and private sphere.

10.6.5 Man-machine interface

In large organisations, the end users of big data analytics are generally not those who implemented the machine learning algorithms or the technological infrastructure that collects, stores and processes data. The data treatment process appears to them as a “black box”. The man-machine interface should be carefully designed to help users in their activities and decision-making, to avoid a “master-slave” relationship between the technical infrastructure, the algorithms and the users. The system should be designed with affordances that help users understand the data treatment process and the underlying assumptions. Additionally, automatic steps enabled by algorithms and software should be carefully designed or limited to enable the user to decide between steps (Blatter and Dechy, 2018). The user should always have a questioning mindset and not blindly accept the outputs of the machine learning process. System designers should also be aware of the ironies of automation (Bainbridge, 1983). Users are not involved in the design and only act when problems arise, or when tasks cannot be automated, and must be handled in unexpected situations without understanding the algorithms.

10.6.6 Incomplete data – formal versus informal, tacit, and quality

Data is (quite obviously!) central to big data analytics, but some key questions concerning quality and relevance of data are sometimes overlooked by analysts in their haste to apply cutting-edge technologies and analysis techniques. Indeed, big data generates dreams of an ideal world in which visits to the shop floor and exposure to the sharp end of operations are no longer necessary, since all relevant information will be pulled into their dashboards. Some important characteristics of sociotechnical system operation such as perceptions, ideas, intentions, beliefs, and non-verbal interactions will remain hard to cover using automated data collection mechanisms.

A fundamental issue to raise is the over-focus on data reported. What about data that is not reported? When some data is collected, what about its relationship to the context in which it was obtained? A lot of context is lost when formalising data and this context must be reintroduced or ‘compensated for’ (Koornneef and Hale, 2004) by the user in their activities and decision-making processes. There are fundamental elements that are not collected: tacit data, informal information, overlooked items. Not collecting them hampers human and organisational factors analysis (especially for root cause analysis). These factors can also dramatically impact quantitative safety analysis. The frequency of occurrence of some conditions and events can be underestimated.

Even the data reported and collected raises questions as well. The data which is collected on system operation is determined and constructed by the worldviews of the people who decide which elements are important to monitor; a narrow worldview may limit the analyses that can be undertaken[55]. This social constructivist perspective of reality should be acknowledged. It implies that even similar data or words, may have different meanings for different people and professional groups. Therefore, critical doubt will remain needed as a complement and check on data codification and automatic decisions based on algorithms.

Data elements are collected, filtered, validated, enriched, analysed and interpreted by multiple people at different phases of the data collection, processing, and decision-making process. The separation between all these people (different professions and objectives) can lead to deviations in the interpretation of the meaning of the data, which can be a source of risk. This phenomenon is particularly relevant concerning data obtained from automated textual analysis, because it is known that different professions or different sites of a same organisation may give different meanings to the same word. In addition, what to report as feedback and as data to collect is affected by political issues within organisations, managerial decisions, and the level of front-line confidence in the reporting system. It could lead individuals and work groups to withhold or under-

[55] *This is an instance of the “What You Look For Is What You Find” or WYLFIWYF problem described by Erik Hollnagel concerning incident investigation.*

report certain events to protect their professional reputation or avoid unwanted intrusions.

Safety analysis is often based on event reports. These reports contain not only textual data, but information and even knowledge from field experts and analysts. Contextual and historical factors, which might be critical to interpreting the textual content of a report, are difficult or impossible to handle using big data and NLP techniques (Rousseau et al, 2018).

Finally, the big data paradigm over-emphasizes the importance of data quantity, and can lead to a "shift from too few to too much data[56]" (Lannoy, 2018). It is known that event recording, especially for near-misses, is not as complete as for serious events, and is sometimes very poor with records registered that are only a few lines for an event in a database. Data quality and the level of coverage of events of interest inevitably impact the level of insight that can be generated concerning safety issues. Efforts for big analytics should be accompanied by renewed investment in the data collection process, not only about its quality, but also about its relevance in particular concerning the human and organisational dimensions of sociotechnical system operation (Dechy and Blatter, 2018; Rousseau et al, 2018).

10.6.7 Data analysis

The streams of data collected by big data infrastructures allow more dynamic analyses than in the past. Data analysis can become more specific and customised with digital twins.

The promises of these new techniques that can find patterns and models in the data may have side effects. Indeed, this added value should not lead to lack of prior analysis, knowledge modelling, formalising heuristics and expert judgment. Real-time thinking for designers and users will require them to have some knowledge readily available; they should not wait for models to emerge from data.

Implementing digitalisation and big data analytics can only work if there is a strong analytical program, with ontological efforts, to clarify rules and models for coding data, language and sense-making issues especially when preparing the machine learning (Rousseau et al, 2018; Dechy and Blatter, 2018).

[56] *https://www.imdr.eu/offres/file_inline_src/818/818_pj_260419_164033.pdf*

In other words, big data analytics should not lead to 'small thinking' (Alloing and Monet, 2016). Rousseau et al. (2018) recalls that the digitalisation and big data issues are not fundamentally new with regard to the questions already raised in the 1980s and 1990s during the first AI and expert system wave.

10.6.8 Machine learning biases

The training data used to build machine learning models may lead to embedded biases which are illegal but difficult to identify. For example, if members of a particular ethnic group tend to have lower than average incomes, they may also have higher rates of incarceration.

Consider an insurance company which builds a machine learning model to estimate credit default risk by feeding input data from existing clients into a large neural network. This neural network may associate specific first names, which are highly correlated with the low-income ethnic group, with higher credit risks, embedding a bias within its decision-support tool that may produce legal problems[57].

10.6.9 Data governance and ethics

We live in a world in which each individual generates 1.7 megabytes of data each second (Petrov, 2020). This rapid data generation brings both opportunities and risks (AIHLEG, 2019). On the one hand, the big data analytics market is estimated to reach $103 billion by 2023 (Petrov, 2020).

On the other hand, there are many evolving risks in our digital landscape such as privacy and security risks (UNDG, 2017; Micheli et al., 2018). The predictive ability of machine learning models may lead to intrusions into people's privacy.

Against this background, data governance is of utmost importance to ensure data is effectively managed and used in an ethical manner. A number of principles for ethical use of big data analytics have been proposed to limit some of these threats (Schwartz, 2010; AIHLEG, 2019).

[57] *See for example* http://www.marketwatch.com/story/big-data-can-lead-to-big-legal-problems-for-companies-2016-06-01 *and [O'Neil 2016].*

Unexpected foresight in retail operations

An annoyed customer walked into a 'Target' store in Minneapolis to complain about the store sending coupons relating to pregnancy products to his high school daughter. A few weeks later, the same customer apologized to the store manager: a discussion with his daughter revealed that she was in fact pregnant [Duhigg 2012]. It is worth noting that an individual's "data footprint" today in 2020 is more than 100 times larger than at the time in 2012.

10.6.10 Invalid conclusions

Appropriate use of machine learning techniques requires high levels of skills in causal reasoning, and subtle mistakes are easily made. Analysis of any large volume of data will very often identify a number of correlations between different variables. Some of these correlations will turn out to be spurious "flukes", and others will be due to the presence of hidden underlying variables, meaning that there is no causal mechanism which could motivate a safety intervention.

Underlying variables

Medical research shows[58] that American men aged between 45 and 82 who skip breakfast have a 27% higher risk of coronary heart disease than other age categories over the 16-year followup period. This does not necessarily imply that eating breakfast reduces heart disease risk; the research also found that people who skip meals may have less healthy lifestyles than average.

In general, it is necessary to implement some form of experiment to check that changing the "predictor" variable does indeed lead to a change in the observed outcome variables. For obvious ethical reasons, this may be difficult to do for safety-related outcomes. The development of a critical view on the validity of inferences made is an important part of training in data analytics.

Invalid conclusion in healthcare

The Cost-Effective HealthCare project analysed emergency room data to try to improve treatment for patients with pneumonia symptoms. They aimed to build a system that could predict people who had a low probability of death, so they could be simply sent home with antibiotics. This would allow care to be focused on the most serious cases, who were likely to suffer complications. The neural network developed by the team had a very high accuracy but, strangely, it always decided to send asthma sufferers home. This conclusion was unexpected, since asthmatics are actually at high risk of complications from pneumonia. It turned out that asthmatics who arrive at the hospital with pneumonia symptoms are always admitted to Intensive Care. Because of this, the training data used to develop the neural network did not include any cases of asthmatics dying, and the model concluded that asthmatics were low risk, when the opposite was actually true. The model was very accurate, but if deployed in production it would certainly have killed people.

10.7 Conclusions

Big data analytics has significant potential to improve the detection of early warning signs of system failure, when compared with the use of standard statistical tools or of unassisted human monitoring of system data. However, while algorithms can improve human judgment, they will never replace it completely. Humans will remain important at every step in designing the process but also in its operation, in (automated or manual) data collection and in expert validation of correlations found. As correlations are not causation, investigation of assumptions made by big data analytics will remain a key activity.

Big data analytics can find patterns in data and derive models from data. This new opportunity does not reduce the importance of traditional analysis techniques, including modelling work to establish a cognitive representation of reality, analyse possible causal links, and extract expert decision-making heuristics. We should take heed from the lessons of the excessive optimism seen during the expert system era of the 1980s and 1990s: implementing big data techniques is not just a technological change that will magically produce results, but must be accompanied

[58] *Prospective Study of Breakfast Eating and Incident Coronary Heart Disease in a Cohort of Male US Health Professionals, Circulation, 2013, DOI: 10.1161/CIRCULATIONAHA.113.001474.*

by critical analysis and expert assessment of the safety relevance of algorithmic predictions.

Effective use of predictive analytics faces a number of obstacles, and requires very specialized skills of the analysts who build the data collection and analysis tools. It also requires a critical viewpoint on the part of users, who will need to assess the relevance of predictions produced by the systems and counter the “black-box” effects of algorithms and integrate contextual factors that may not have been taken into account. The man-machine interface is of critical importance to allow step-by-step control of the data analysis process, back and forth, in a master-slave relationship. Automatic decision-making based on algorithms should be very carefully controlled due to many biases in quality of data (including under-reporting) and its treatment, the difficulty of handling data on human and organisational factors of system performance, which are often tacit and informal.

10.8 References

Alloing, C., Moinet N. (2016), *Les signaux faibles, du mythe à la mystification*, Hermès, La Revue, 2016/3, n°76, pp. 86-92.

Amanullah, M. A., Habeeb, A. S. M., Nasaruddin, F. H., Gani, A., Ahmed, E., Nainar, A. S. M., Akim, N. M., Imran, M. (2015). *Deep learning and big data technologies for IoT security*, Computer Communications, Volume 151, 1 February 2020, Pages 495-517, DOI : 10.1016/j.comcom.2020.01.016

Antonsen, S., (2009). *Safety Culture Assessment: A Mission Impossible?,* Journal of Contingencies and Crisis Management 17:4, DOI: 10.1111/j.1468-5973.2009.00585.x.

Artificial Intelligence High Level Expert Group – AIHLEG (2019), Ethics guidelines for trustworthy AI. https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai

Bainbridge, L. (1983). *Ironies of automation*. Automatica, 19, issue 9, 775-779.

Baker J., Bowman F., Erwin G., Gorton S., Hendershot D., Leveson N., Priest S., Rosenthal I., Tebo P., Wiegmann D., Wilson L. (2007), The Report of the BP U.S. Refineries Independent Safety Review Panel.

Dechy N., Blatter C. (2018), *Better addressing human and organisational factors in learning from experience: what benefits and challenges for big data and automatic language processing?* (in French) in Proceedings of the Institut pour la Maîtrise des Risques λμ21 conference, Reims, October 2018.

Duhigg, C. (2012). *How companies learn your secrets.* New York Times.

Espinosa, A. J., Kaisler, S., Armour, F. And Money, W.h. (2019). Big Data Redux: New Issues and Challenges Moving Forward, Proceedings of the 52nd Hawaii International Conference on System Sciences, HICSS 52, Maui, Hawaii, 2019. https://scholarspace.manoa.hawaii.edu/bitstream/10125/59546/0106.pdf

Esteva A., et al. (2017) *Dermatologist-level classification of skin cancer with deep neural networks,* Nature 2017; 542: 115-118.

Hagras H., (2018) *Toward Human-Understandable, Explainable AI* in Computer, vol. 51, no. 09, pp. 28-36, 2018. DOI: 10.1109/MC.2018.3620965

Ham, D.-H., Park, J. (2020). *Use of a big data analysis technique for extracting HRA data from event investigation reports based on the Safety-II concept*, Reliability Engineering & System Safety, Volume 194, February 2020, 106232, DOI: 10.1016/j.ress.2018.07.033

Huang, L., Wu, C., Wang, B. (2019). *Challenges, opportunities and paradigm of applying big data to production safety management: From a theoretical perspective*, Journal of Cleaner Production, DOI: 10.1016/j.jclepro.2019.05.245

Jouniaux P., Hadida D., Dechy N., Marle L., Billy F., Pierlot S., Parrennes F., Rouvière G., Husson D. (2014), *Detection, relevance and amplification of weak signals within the learning from experience*, (in French), in proceedings of the Institut pour la Maîtrise des Risques λμ19 conference, Dijon, October 2014

Khan, M. A., Uddin, M. F. and Gupta, N. (2014), *Seven V's of Big Data - Understanding Big Data to extract Value*, Proceedings of 2014 Zone 1 Conference of the American Society for Engineering Education (ASEE Zone 1), http://www.asee.org/documents/zones/zone1/2014/Professional/PDFs/113.pdf, ©2014 IEEE

Koornneef, F., and Hale, A. (2004), Organizational memory for learning from operational surprises: requirements and pitfalls, in Andriessen J. H., Fahlbruch B., How to manage experience sharing – From organisational surprises to organisational knowledge, Elsevier science, Amsterdam

Micheli M., Blakemore M., Ponti M. and Craglia M. (Eds.) (2018*), The Governance of Data in a Digitally Transformed European Society*, Second Workshop of the DigiTranScope Project, JRC114711, European Union, 2018. https://ec.europa.eu/jrc/communities/sites/jrccties/files/jrc_digitranscope_report_-_oct_2018_data_governance_workshop_1.pdf

O'Neil, C., (2016), *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, ISBN: 978-0553418811.

Petrov, C. (2020), 25+ Impressive Big Data Statistics for 2020, Techjury Blog. https://techjury.net/blog/big-data-statistics/#gref

Qiao, W., Liu, Q., Li, X., Luo, X. and Wan, Y. (2018), *Using data mining techniques to analyse the influencing factor of unsafe behaviors in Chinese underground coal mines*, Resources Policy, Volume 59, December 2018, Pages 210-216, DOI: 10.1016/j.resourpol.2018.07.003

Rousseau, J.-M., Montméat A., Ben Ayadi M., Hébraud C. (2018), *Operating experience feedback and digitalisation – No pain, no big data!* (in French) in Proceedings of the Institut pour la Maîtrise des Risques λµ21 conference, Reims, October 2018

Rowley, J. (2007), *The wisdom hierarchy: representations of the DIKW hierarchy*, Journal of Information Science, 33(2) 2007, pp. 163–180, DOI: 10.1177/0165551506070706

Scheerlinck, J, Van Eeghem, F. and Loutas, N. (2018), D05.02 Big Data Interoperability Analysis, Report written as part of Action 2016.07, SEMIC (Semantic Interoperability Community) of the Interoperability solutions for public administrations, businesses and citizens (ISA²) Programme, Deliverable for task 5 - Capacity Building on Information Governance, Specific contract 508. https://joinup.ec.europa.eu/collection/semantic-interoperability-community-semic/document/big-data-interoperability-analysis

Schwartz, Paul. M., (2010) *Data Protection Law and the Ethical Use of Analytics*. The Centre for Information Policy Leadership, Hunton & Williams LLP

Silver, N., (2012). *The Signal and the Noise: Why Most Predictions Fail – but Some Don't*. Penguin Books. ISBN: 978-1594204111.

World Economic Forum – WEF (2019), *How much data is generated each day?* https://www.weforum.org/agenda/2019/04/how-much-data-is-generated-each-day-cf4bddf29f/

This chapter is extracted from the final technical report of the ESReDA Project Group *Foresight in Safety*. The full report is freely downloadable from the ESReDA web site and from the EU Joint Research Centre publications repository.

Bibliographic identifiers for the full report are indicated below.

PDF	ISBN 978-92-76-25189-7	ISSN 1831-9424	doi: 10.2760/814452	KJ-NA-30441-EN-N
Print	ISBN 978-92-76-25188-0	ISSN 1018-5593	doi: 10.2760/382517	KJ-NA-30441-EN-C

“Enhancing Safety: The Challenge of Foresight”

Edited by the ESReDA Project Group *Foresight in Safety*.

How to **cite this report**: ESReDA Project Group Foresight in Safety, *Enhancing Safety: The Challenge of Foresight*, EUR 30441 EN, Publications Office of the European Union, Luxembourg, 2020. ISBN 978-92-76-25189-7, doi: 10.2760/814452, JRC122252.

ESReDA
European Safety, Reliability & Data Association