Safety and Health at Work 12 (2021) 1–9

Contents lists available at ScienceDirect

Safety and Health at Work

journal homepage: www.e-shaw.net

Review Article

# Finding Pluto: An Analytics-Based Approach to Safety Data Ecosystems

Thomas T. Barker*

*The University of Alberta, Faculty of Extension, Canada*

ARTICLE INFO

*Article history:*
Received 2 July 2020
Received in revised form
14 September 2020
Accepted 21 September 2020
Available online 1 October 2020

*Keywords:*
big data
data science
safety

ABSTRACT

This review article addresses the role of safety professionals in the diffusion strategies for predictive analytics for safety performance. The article explores the models, definitions, roles, and relationships of safety professionals in knowledge application, access, management, and leadership in safety analytics. The article addresses challenges safety professionals face when integrating safety analytics in organizational settings in four operations areas: application, technology, management, and strategy. A review of existing conventional safety data sources (safety data, internal data, external data, and context data) is briefly summarized as a baseline. For each of these data sources, the article points out how emerging analytic data sources (such as Industry 4.0 and the Internet of Things) broaden and challenge the scope of work and operational roles throughout an organization. In doing so, the article defines four perspectives on the integration of predictive analytics into organizational safety practice: the programmatic perspective, the technological perspective, the sociocultural perspective, and knowledge-organization perspective. The article posits a four-level, organizational knowledge-skills-abilities matrix for analytics integration, indicating key organizational capacities needed for each area. The work shows the benefits of organizational alignment, clear stakeholder categorization, and the ability to predict future safety performance.



## 1. Introduction

Clyde Tombaugh, who discovered the ninth planet in the solar system—the cold and distant object now known as Pluto—made the discovery by going through an interesting process of detective work. In fact, the mysterious existence of this unobservable planet was based on observable astronomical data: wobbles in the orbits of Uranus and Neptune [1]. The existence of the object was proposed in theory as a possibility or a "missing" object before it ever showed in a telescope. For those interested in the role of Occupational Health and Safety (OHS) professionals in safety analytics, Tombaugh's work models a shift of thinking: from identifying observed objects to postulating objects based on data from other observations and other sources, some only tangentially related to the object being sought. OHS professionals face a similar shift of thinking. If we imagine "safety performance" as the focus of attention (identifying a behavior) for OHS professionals, then we can imagine an environment (or ecosystem) in which measures other than direct observation (incident investigation) might lead to "finding Pluto." The correctly *postulated* safety behavior is the "Pluto" of analytic OHS work with big data systems. Finding "Plutos" suggests the possibility of finding safety incidents and performance profiles.

Tombaugh's direction was forward-looking (predictive) rather than backward-looking (quantitative). Similarly, OHS systems have traditionally focused on past events; generalizing about incidents and using the data to design mitigation systems. But this backward-looking has not allowed them to predict incidents or identify safety behaviors with high degrees of certainty [2]. The goal of OHS predictive data analytics, also known as *safety informatics* or *safety intelligence,* is not to quantify patterns of errors and system failures, but to sense out hard-to-find *archetypes* of safety performance [3–6]. To reach new levels of effectiveness, OHS professionals need to start "finding Plutos." This kind of predictive analytical thinking, or the tracing of "weak and potential information" opens opportunities to mine the knowledge hidden in *big data systems*—a catch-all phrase here for "known systems"—such as we will explore later in this paper [6]. Theoretically, these big data systems—rightly identified, accessed, managed, and capitalized—hold the keys to finding the "missing planets" in the solar system of OHS [7].

* Corresponding author. 2-367 Enterprise Square 10230 Jasper Avenue NW, Edmonton, Alberta, T5J4P6, Canada.
*E-mail address:* ttbarker@ualberta.ca (T.T. Barker).

2093-7911/$ – see front matter 
https://doi.org/10.1016/j.shaw.2020.09.010

**Table 1**
Variables for predicting safety performance

| Variable | Applicability to safety analytics |
| --- | --- |
| Employment security | Employment security has been linked with safety performance. In addition, low performance has been linked to lower motivation and low safety compliance. |
| Selective hiring | The fit of employees for their job has been shown to correlate with positive safety performance. Employees with a better job fit have better chances of avoiding injuries. |
| Extensive training | The link between safety and performance shows the impact of safety training on employee engagement. According to Zacharatos, “organizational commitment predicts work performance in general and safety performance in particular.” [18] |
| Self-managed teams and decentralized decision making | Team-working variables can indicate cohesion and also help safety analysts find instances of shared responsibility and accountability for making positive safety decisions. |
| Status distinctions | The more an organization draws lines between “front-line employees” and those in other levels of employees, the more a “blame” framework develops. Reducing status distinctions may predict safety performance. |
| Information sharing | Sharing safety information can lead to safety performance. “Organizations with better safety programs were characterized by more open discussion between management and employees.” [18] |
| Compensation contingent on safety performance | Data indicates that safety performance is improved when employees are convinced that their work is valued and rewarded. |
| Transformational leadership | As a variable, transformational leadership can indicate key to high safety performance and is associated with greater safety and lower job injuries. |
| High-quality work | Interesting, meaningful work can be a positive predictor of safety performance. |
| Measurement of management processes | Factors relating to effective management processes can predict subsequent safety performance. |
| A system of high-performance practices | Safety performance systems benefit from an integrated approach that recognizes the interdependence of organizational systems. |

This article is a theoretical and practical exploration that sets the stage for transformed OHS professional practice by looking first at the variety of data sources available to OHS ecosystems, and then at the practical guidance provided by professional role descriptions [8,9]. Following that stage-setting, the article will present a conceptual system for envisioning the integration of data analytics into an organization. This integrated structure of roles and responsibilities builds on the scholarship of Wick in identifying the diffusion of knowledge and knowledge management in organizations [10]. In that earlier work, Wick focused on the roles of the technical communicator, the IT specialist, the manager, and the executive as each having a key role and perspective on knowledge integration. This paper extends that thinking by postulating a similar role of integration in terms of the OHS professionals and predictive data analytics.

It is important to remember that the claims for OHS effectiveness at the level of these imagined ecosystem models need to be based on processing capabilities that go beyond, or extend, traditional channels of information and analytical skills heretofore available to many professionals, not just those in safety; and that go beyond the isolated safety project or initiative [11]. The challenge is to connect the dots so that positive safety outcomes not only persist, but improve [12].

To address the role of safety professionals in the diffusion strategies for predictive analytics tools, the following two questions must be answered:

1. How has the work of safety professionals changed in light of predictive analytics and professional requirements?
2. How can predictive analytics be implemented in programming, technological systems, management and strategic leadership?

To address these questions, we will (1) briefly survey the sources of predictive analytics data available to OHS professionals; (2) examine four perspectives on predictive analytics based on these investigations; and (3) conclude by positing a four-level knowledge-skills-abilities model for organizational analytics integration, indicating key capacities needed for each operational unit. The results of this work will yield insights into new leadership opportunities for OHS professionals.

## 2. Background for shifts in safety thinking

Two things—(1) developments in data availability and the information technology to exploit it, and (2) revisions to OHS certification requirements—have focused the attention of safety professionals in North America and globally on new data sources that may improve health and safety outcomes for business and industry. The recent developments in data availability, under the journal categories of psychology, management, and technology, are known by various terms: *big data, digital ecosystems, advanced informatics, predictive analytics, integrated digital systems,* and *data analytics* [13]. These emerging, multidisciplinary sources of data promise to supplement or replace traditional historical incident data, and also to transform enterprise thinking about the value added by advanced health and safety programs [14,15].

One way to look at big data is through its “v” characteristics, defined as volume, velocity, variety, value, variability, and visualization [14,16,17]. A summary of these characteristics shows that, in addition to high-volume sources like social media and The Internet of Things, big data sources are being applied to a wide variety of disciplines, as Li's 2016 review indicates [15]. A key characteristic of big data is that its analysis and application (strategically and operationally) apply across sectors and disciplines [14]. One way to see how it applies specifically to safety and safety performance is in the work of Zacharatos et al., who review a number of applications of big data to safety, and who identify the applicability of data that is mined from high-performance safety systems [18,19]. The argument made in this review is that, “high-performance work systems can be applied to improving workplace safety just as well as economic performance.” (p. 78) Li et al., also show how *biometrics* is emerging as a way to visualize these broader research domains [20]. This assertion aligns with the view of safety as a “performance variable,” making the target of predictive explorations not safety incidents but, more broadly, the *conditions for positive safety performance* [8,21–24]. Table 1 offers a summary of these high-performance variables, with indications of how they might be useful in predicting safety performance.

The second motivation to improve health and safety outcomes comes from new professional certification requirements. Recent revisions of the International Network of Safety and Health Practitioner Organizations (INSHPO) and the Board of Canadian Registered Safety Professionals certification requirements for OHS professionals have redefined the role of data analytics in all aspects of the safety enterprise: technological and process tools, program design, safety project management, and strategic planning for safety culture growth and competitive advantage in knowledge-rich industries. For example, The *Professional Capability Framework: A Global Framework for Practice*, published by INSHPO in October, 2015, states that the *Professional* role in OHS requires critical thinking, information gathering, communication skills, and judgement to identify and analyze complex OHS problems to generate practical evidence-based solutions." (p. 8) Although some work on project-level data integration has been done [11], just how *organizations* will shift to new sources of data and shift the thinking about safety initiatives remains to be sorted out.

### 2.1. The workplace safety ecosystems model

This organizational shift can be seen as a shift from an *individualized safety model* (with data collection and analysis focusing on past incidents) to a focus on a *workplace safety ecosystems model*, with data collection and analysis that draws on a variety of data sources, including huge data repositories [6,25–27]. Some researchers refer to these two ways of thinking as "Safety I" and "Safety II." [9,28] One such data repository, for example, is the data accumulated by smart devices and machines: the so-called *Internet of Things* [29]. The emerging field of *safety informatics* and "smart safety decision-making" are revolutionizing the role of big data (or "big safety data") in safety thinking [3,5,30]. Another source of information, for example, is the "open web." [31] Dasgupta claims that, "curating and managing health and safety related incidents from various sources is an important part of an organization." (p. 434) Still another source is the paradigm shift occurring in OHS and safety known as Industry 4.0 [32]. Industry 4.0 represents the convergence of manufacturing with the digital revolution, artificial intelligence, the Internet of things, and every device called "smart" [32]. According to Badri et al. "In Industry 4.0, businesses digitize their physical assets and integrate them into digital ecosystems throughout the value chain." [32] As a result of this digitization of manufacturing, these sources of data become the building blocks for safety ecosystems models, rather than individualized safety models.

## 3. Sources of data for OHS analytics ecosystems

According to scholarly reviews of safety analytics, the clues for finding objects or incidents in data not collected historically or in incident records, can come from hidden variables: variables "not captured in incident reports." [23,33] These types of data include: equipment operation and process data [34,35], vehicle telemetry [36,37], weather, geo-spacial, socio-demographic, human resources (payroll, performance data) [38] and training [39], industry and other data. For example, author networks using biometrics are important ways to visualize big-data resulting from meta-analysis [34]. These data also include new developments in Industry 4.0, the Internet of Things, artificial intelligence, enterprise wearables, and other environmental safety reporting systems. In the following section, we will focus on four categories of information as a way of organizing new and existing sources, and making sense of them. The categories chosen here as broad categories of *existing sources* of data are: safety data, HR data, context setting data, and external data [33]. These categories provide an overview of the range of data sets available and will suffice for the expository task at hand.

### 3.1. Safety data

Safety data consist of the more conventional, historical data collected by government and international organizations and used by OHS professionals [40]. Few doubt that these sources of information, well analyzed, can help predict and prevent future incidents. Technically, such data are not new to OHS. What is new is the analytical integration of them into accident mitigation strategies based on predictive analytics. Doing so requires an advanced degree of integration. One such type of safety data available to OHS professionals is that of *governmental regulatory groups*, who gather and maintain statistical record systems [41]. In Canada, for example, such data would be compiled by the Canadian Centre for Occupational Health and Safety [42]. Information available from this source includes data collected in the National Work Injuries Statistics Program and is supplied by workers compensation boards across Canada [43].

These conventional sources of safety data, often consulted by OHS professionals, are summarized in the textbook *Occupational Health & Safety: Theory, Strategy & Industry Practice* [44]. Conventional sources include data collected through social-science methods: interviews, focus groups, surveys and questionnaires, observations, documentation, and laboratory experimentation [44]. As with all data sources, the challenge, as Dyck notes, is to avoid mental filters and allow a professional to "see an issue from many perspectives." (p. 1191) *Audits* provide another source of data, characterized by an attempt to measure results "against a predetermined protocol." (p. 118). Audits are excellent sources of information about the success of OHS programs and management systems. Similarly, *incident investigation* is a standard source of data. Again, in the spirit of effective use of incident investigation data, researchers suggest it "should be retrospective and prospective simultaneously." [45].

The point of reviewing these conventional sources of safety data is that big-data sources suggest innovative data analysis frameworks. For example, a key to understanding and applying existing safety data systems is to identify and focus on *lagging indicators* [46]. However, as Hollnagel reminds us, "for effective safety management in general, it is necessary to have *leading indicators*." Leading indicators are not the only way to address the challenge of predicting incidents: variables needed in the analysis of big-data systems should account for the variability of human detection, observation, and categorization [9,45].

In some ways, conventional sources of information gathering, analysis, and action are similar to or extensions of existing systems. For example:

- Root cause analysis gives way to machine learning algorithms [22,29,45]
- Observational data gives way to big-data techniques [22]
- Historical data gives way to real-time "data-fusion frameworks" [22].

As these observations suggest, the goal of safety analytics is often not just to know what happened, and to whom, but to hunt in between these known perspectives to find new correlations. For example, the safety data set in the offshore drilling industry today is immeasurably larger than it was in previous years, because data are not only being collected in research events, but *moment by moment* thanks to safety wearables, cameras, sensors, and robotics reporting. Concepts like "association rule" become important, because they allow the safety professional, safety manager, safety software engineer, or safety strategist to make sense by visualizing categories, many of which may be suggested by artificial-intelligence algorithms. Trevistan et al., make this point in looking at data-mined information about offshore safety performance factors [47]. Visualizations, such as the one in figure are [47], illustrate the way relational concepts portray pathways through data sets, pathways that can lead to understandings of clusters of factors that model incidents of safety performance. In figure are [47], the slider at the top allows for cycling through "association rules" to discover patterns in "antecedents" and "consequences" (roughly equivalent to conditions *before* and *after* an incident). These associations show up as darkened areas in the visualizations. Savvy safety professionals can use these associations to identify potential safety behaviors, unforeseen in the broader context.

## 4. Context-setting data

What is the *broader context* of any given safety incident or safety performance event? Answers to this question can often be found in conventional sources of risk management information: What task was being performed? What site variables were at play? What equipment was being used? How complex was the resource and production environment? These variables record the shape of safety performance [44]. Such a *positive* view of the context has not always been the case. As Dekker points out, in the past, the negative emphasis was on human errors and accidents rather than on positive safety performance [8,48]. The positive emphasis on safety performance, however, which is characteristic of the *Safety Differently* approach, suggests that expanded data sets of contextual information can reinforce the motivation to build on safety success rather than punish for safety failure or non-compliance.

To that positive end, a number of innovative trends in analytics take advantage of big data to search in unusual places to discover emergent contexts of safety performance. These trends include: robotics and cobotics [49,50], semantic photography [51], and sociometric visualization [52,53]. The expansion of data sources that we see in analytics work means that what might have appeared as an anomaly in, say, incident data, can be correlated (e.g., using a cross-entropy estimation approach) that aligns with historical data from other sources and helps shape information found in incident and contextual data [54]. The difference with big data is that these shifts of knowledge emphasis can happen, as Bifet points out, in real time.

Streaming data analysis in real time is becoming the fastest and most efficient way to obtain useful knowledge from what is happening now, allowing organizations to react quickly when problems appear or to detect new trends helping to improve their performance [55].

### 4.1. Internal data

Conventional sources of organizational data used in safety planning rely heavily on the shaping of incident targets based on human resources operational units and their information systems. Dibenedetto identifies a number of conventional sources of data to inform existing OHS information management systems. These sources include: exposure monitoring records, Workers Compensation Board records, equipment calibration and maintenance logs, motor vehicle records, vendor lists, as well as chemical inventories and material safety data sheets [56,57]. Much of this kind of information can be obtained through a healthy interaction with human resources information systems. Rosters and staff allocation systems that schedule and manage safety performance interactions provide data collected through staffing and scheduling enterprise software [58]. Additional sources of resource operations data include performance histories, skill and training evaluation information repositories, all which make up conventional internal sources of OHS data. However, a key distinction to make between new and conventional sources of operational information—which is what we're talking about here—lies in the sheer size of the data sets, and advances in data-mining methods and algorithms to use in these humongoid human-resources data sets. Neural-network data, for example, provides vast data sets that can inform the prediction of job and safety performance [59,60].

But perhaps the most important trend in big-data analytics is the swapping of a *method-driven* approach (that looked at specific OHS variables) to a *domain-driven* approach that takes an integrated and multidimensional (*data ecosystems)* approach. This approach is illustrated in systems that evaluate targeted organizational business processes with the aim of optimizing all processes. Morabito illustrates this in systems like Invenio™, that evaluate "application log files, documents, email messages, and social signals to optimize an organization's business processes." [61] The advantage of this approach is the discovery of "underlying" company processes. Such company processes, seen as ontological or formative structuration elements, can reveal archetypes of safety performance stemming from unforeseen causes in unforeseen settings. Other systems reviewed by Morabito offer similar innovative ways of looking at internal, domain-specific data.

### 4.2. External data

External data sources of organizational and broader contextual influences—stakeholder benchmarks, cultural elements of the social and political (and other) economies, and sociodemographic data—provide important clues as to safety performance, both from regressive and proactive viewpoints. Often these considerations come under the umbrella term "diversity." [44] Unhappily, cultural assessment is often unstructured or hap-hazard, or worse, based on inherent or implicit bias. Also, cultural assessments, including sociodemographic variables, are often carried out by surveys that target known characteristics and outcomes [62]. The question for OHS strategists lies in how to augment known qualitative measures with "new techniques for automated analysis of large amounts of text in iterative fashion." [63] Some suggestions for handling external data include: using data framing for text [63,64], using geo-coding for location tagging [64], and "knowledge network analysis" for initial processing and high-level modelling of safety performance [64].

To summarize, conventional sources of data available to safety professionals took the form of safety (incident) data, context-setting data, internal data, and external data. The availability of and ability to process much larger data sets from tangentially related organizational, material, and governmental systems has expanded the purview of the safety professional. Both that purview and the methods of exploiting it have been enhanced by data analytics. The aim of understanding in this area is to derive some of the potential benefits of safety analytics.

### 4.3. The benefits of advanced data analytics in OHS

The following sections provide an overview of new perspectives in OHS through the lens of expanded data sources. As the earlier discussion illustrates many conventional evidentiary sources of data, including targeting and reflective methods like surveys and interviews, as well as conventional environmental and industrial

**Table 2**
The benefits of advanced data analytics in OHS

| | |
|---|---|
| 1. | **Real-time ecosystem monitoring.** Much of the direction and results of big data analytics focuses on creating systems that mine, process, and present data from integrated sources *in real time.* Thus, the ability to vary the query or focus promises to enhance the exploration of the factors of safety performance. |
| 2. | **Stakeholder segmentation.** As a core function of health and safety practice, stakeholder segmentation is enhanced [15,27]. |
| 3. | **Descriptive to predictive.** The emphasis on predictive power holds great promise of mitigating personal and environmental factors to enhance safety performance. Nuanced examination of descriptive behaviors that ranges into uncharted data territory holds the promise of filling safety performance gaps. |
| 4. | **Follows leading indicators.** A crucial distinction between previous ways of seeing OHS planning is that big data analytics requires analysis of leading indicators of ongoing streams of data collected for various reasons and not specifically for OHS purposes [21,22,49]. |
| 5. | **Bottom-up, domain-driven framework.** [57] This shift in approach pays off in producing insights about safety performance that are "bottom-up" and have thus enhanced possibilities in filling safety performance gaps. |
| 6. | **Relevance over method.** If for no other reason than flexibility and a sense of adventure, what actually works to better assist in perceiving safety performance will work to create clearer categories than alignment with existing, and often siloed, domain knowledge (and their information systems) [57,65,73]. |
| 7. | **Big is better.** Some data representations are suitable to limited, historical, and descriptive data sets (decision trees) while other representations, such as neural networks that mimic biological neural networks require huge amounts of information. Thus, big data sets and systems allow for data representations that are conducive to explorations into future safety performance [27,58,65]. |

demographics, provide the direction of safety program development beyond what has guided practitioners in the past. The preceding brief overview of big data analytic perspectives, built to explore greatly expanded data sets using mining and other methods, offers insights into the benefits of modelling safety performance in this way. These benefits are summarized in Table 2.

According to Huang et al., "Access to large-scale, fast-moving, complex streams of safety data-sets has the potential to fundamentally transform the way organizations make their safety decisions. Accordingly, the [big data] support for [safety decision-making] at all levels of organizations, along with the way they are organized, is becoming increasingly critical." [16] In the following section, we provide a prototype matrix of perspectives on this support, and derive some characteristics of that transformation.

## 5. Perspectives on predictive analytics in safety

The preceding discussion of sources that facilitate data analytic methods supports two important points. First, innovative analytics methods can be seen as *enhancements to existing and conventional sources*, rather than as entirely new sources. Second, for the OHS professional, the disparate viewpoints on data analysis and data-driven predictive analytics may best be seen as the functioning of an entire *data ecosystem*—an amalgam of data sets, analytical algorithms, visualization and knowledge representation systems—all guided by the motivation to find the ever-illusive, unseen model of safety performance. However, as Ouyang et al. remind us, the question remains: How can these valuable approaches to analytics be integrated into an organization? [6].

Scholarly literature on the integration of analytics into organizations suggests two main integration strategies: centralized and decentralized. Grossman and Siegel address this question, noting that " … there is as yet no consensus about how best to organize analytics efforts within the organization and what core analytics processes the organization must support." [65] The authors review three basic models of how to integrate analytics functions in an organization: (1) centralized analytics; (2) decentralized analytics; and (3) a hybrid model, whereby a critical mass of data scientists are housed in a central unit, but also collaborate with data scientists distributed throughout the organization [65]. However, Davenport approaches analytics integration from the perspective of competitive advantage, and stresses that 11 of the 32 organizations they surveyed that had robust data analytics initiatives, "managed analytical activity at the enterprise (not departmental) level." [66] Pence supports this enterprise-level view, pointing out that an enterprise-wide integration effort is required to clearly identify and respond to "underlying organizational mechanisms" and their information systems that might support predictive analytics objectives [64]. Still another view is taken by Ouyang et al., who look at big safety data integration from a three-part perspective: the data set, perspective, the technology perspective, and the safety thinking or leadership perspective [6].

In this article, we take the enterprise-wide view, going boldly where safety professionals have not gone before. The question these models leave us with, however, is, How can we view the integration of analytics into organizations in a way that encompasses (and facilitates) the safety perspectives of core programmatic, technological, managerial, and strategic units that are more or less ubiquitous among organizations seeking to implement analytics?

The next section, I will outline such an integration model that attempts to account for four organizational perspectives on the big-data analytic activities: the program perspective, the technological perspective, the socio-organizational perspective, and the knowledge-organization perspective. The people behind all these perspectives face decisions and follow processes suitable for making knowledge from observed and recorded behavior. In some work circumstances all of the roles below may be carried out by one individual; but, as we shall see, specialized professional domain knowledge is needed to shape the overall use of safety analytics. As a form of ontology of safety analytics, these perspectives provide "a foundation for reasoning about the domain." [22].

## 6. The programmatic perspective

The programmatic perspective on predictive solutions has to do with integrating the solution into the existing OHS management programs, OHS policies, and program review mechanisms [44]. To accomplish this the practitioner needs to address content, training, evaluation, and continuous improvement of safety programs, as we saw earlier, within the constraints of layers of national and international standards [24]. In doing so, the training and intervention designers need to clearly link to leading indicators of safety performance [44]. They inhabit the world of front-line safety performers and their job is implementation. The "data-buck" stops with safety programmers.

The following four points summarize the programmatic perspective.

- **Expertise**: The work requires prioritization of interventions, message-mapping and documentation, campaigns and

storytelling, monitoring and evaluation, and review and recursive knowledge storage [41].
- **Challenge**: The challenge is in representing the model in programs: how well does the Pluto that we have found match the Pluto in the mind of the risk stakeholder, for instance [11,41].
- **Process**: The process requires constructive steps to implement analytics models. The process requires a “holistic approach to complex project planning” [11].
- **Tasks**: The task is programming forward-looking and predictive analytical solutions [67].

## 7. The technological perspective

The technological perspective on predictive analytics encompasses the domain of information technology and software systems design. In regard to the five Vs of big data, this perspective focuses often on the *velocity* of data and attempts to process data in real or near-real time frames [29,68]. Unlike the programming emphasis in the previous perspective, the technological perspective is often associated with the processing stages of big data: development of algorithms, establishing statistical validity, and providing knowledge representations appropriate for operational and strategic stakeholders [69]. Professionals in this role often message to internal clients based on business and data understandings, data preparation and modeling, as well as evaluation and deployment of knowledge models [29]. Of concern to IT professionals are challenges in the design of database file systems, knowledge base structure, and search and analysis algorithms. They inhabit the world of vendors and solutions.

The following four points summarize the technological perspective:

- **Expertise**: The perspective requires highly specialized expertise in system architectures or approaches to processing and analysis [29].
- **Challenge**: The challenge lies in effective extraction (validating, cleaning, recording, and transmission) from sources collected for a purpose other than safety [70].
- **Process**: The process revolves around data-mining objectives and model development [70].
- **Tasks**: The data representation and visualization tasks must accommodate data from text, natural language processing, and internet safety articles [31,70].

## 8. The socio-organizational perspective

The management of analytics initiatives—the people side of big-data thinking—is the purview of the socio-organizational perspective. The emphasis in the socio-organizational perspective, as the name implies, is in the social and cultural nature of safety knowledge. The goal of this perspective is the integration of the work done by IT professionals and safety training specialists into the work of organizational stakeholders to achieve goals of modularity, reusability, and scalability. They inhabit the world of management and system integration.

The following four points summarize the socio-organizational perspective:

- **Expertise**: Managers need expertise in prediction of safety incidents based on workload analysis, skill requirements calculations, problem formation, operator adjustment, and other factors [71,72].
- **Challenge**: A primary challenge is the building of safety culture [8,21,29]. This challenge requires “the marriage of cultural sociology and big data” [63].
- **Process**: This perspective reflects the ontological impulse that not only *defines* the organization as a data ecosystem, but *constitutes* it [64,73,74].
- **Tasks**: The key task is integrating various strategic perspectives (internal cohesion, ecosystem identity, knowledge work flows, resource sharing protocols, data access principles, and repository sharing protocols, need analyses, and evaluative and reporting channels) on big data to facilitate OHS strategic goals [58,73,74].

## 9. The knowledge organization perspective

The knowledge organization perspective on big-data analytics emphasizes the building of knowledge capital from internal systems. Taking the broad and industry-wide view, the goal is to set strategy for an entire safety ecosystem. Messaging from this level of analytics should be to internal groups, such as stockholders, but primarily to external industry stakeholders: professional organizations, industry advocacy groups, financial organizations, government, and legal systems. Ideally knowledge organization addresses issues that arise in competitive arenas. Knowledge organization takes the lead.

The following four points summarize the knowledge organization perspective:

- **Expertise**: This evolution of strategic industrial capability (as in Industry 4.0) sees and values digital processes, artificial intelligence, data from smart devices, and the Internet of things as central to future efficiency [32,75]**.** The key is to build safety knowledge capital.
- **Challenge**: The challenge is to discover evidence that supports the success for organizations in using data analytics to extract valuable insights from big data that support strategic decision-making [16].
- **Process**: Knowledge leaders must assess knowledge capital gains from process initiatives and then communicate them to external strategic stakeholders.
- **Tasks**: Organization knowledge leaders use communication channels to articulate key strategic transactions imbricated in enterprise-wide systems.

These four perspectives on data analytics and OHS form a continuum from *data* to *data application* to *data culture* to *data value*. The data are identified by IT professionals, cleaned and made available to practitioners, who apply it in programs. Managers align the use of the knowledge to other operational units, and, over time, executives shape instance after instance into intellectual capital and strategic directions. But it may not work quite this simply. To be really effective, analytics implementations need to be *integrated*. Warp speed needs to have ripples.

### *9.1. An integrated OHS big data analytics model*

Scholars on the diffusion of innovations such as enterprise-wide big-data analytics often point out that integrated analytics activities require an exploratory and almost self-generating direction [61,64,66]. This exploratory direction means that by working with open systems in a knowledge organization, all purposes contribute to the acquisition and processing of meta-analytic data that further energizes efficiency and production. In this somewhat ideal scenario, a hypothesis, an innovation, an idea, a hunch (be it made by a human, robot, or device) can trigger simultaneous ripples through the big-data analytics ecosystem, leading to the productive conclusion of hypothesized and discovered instances of

**Table 3**
Knowledge, skills, and attributes associated with perspectives on predictive analytics

| | Programmatic perspective | Technological perspective | Socio-organizational perspective | Knowledge organization perspective |
|---|---|---|---|---|
| **Knowledge** | Deep domain knowledge of hazards and risks<br>Correlations between predictions and current safety performance | Operating systems, platforms, applications<br>Statistical modeling and algorithm theories | Theories of diffusion of technology, culture, organizational integration<br>Organizational models for predictive analytics | Broad industry and market trends and opportunities<br>Indicators of knowledge capital |
| **Skills** | Design, develop, and evaluate safety programs<br>Engagement with safety stakeholder perspectives | Design, develop, visualize, support, and evaluate distributed systems<br>Implementation of analysis models systematically | Communicate and manage integrated systems<br>Change management | Strategic planning<br>Transformational leadership |
| **Abilities** | Effectively articulate safety performance behaviors | Create hardware and software integration across units<br>Design and program data analytics interfaces using knowledge-representation schemas | Get everybody working from the same playbook | Lead the organization based on alignment to core values |

safety performance. If the big-data analytics framework is a true *information ecosystem*, then the sections of it will articulate in innovative ways.

Just how the innovation of data analytics in OHS will shape and be shaped by functional units in an organization is the next direction of this exploratory article. In this final section, I posit a four-level knowledge-skills-abilities model for analytics integration, indicating the key capacities needed for each level. It is important to realize that all the levels share the ability to interact with and integrate with the other three perspectives, with the goal of growing the data ecosystem. Table 3 summarizes the four perspectives.

The chart in Table 3 represents a solar system of analytic knowledge about safety performance. Although the many challenges of implementing big-data analytics that it articulates have to do with database creation, data sources, visualization and exploration, and predicting safety models for injuries, accidents, and illnesses, equally important are considerations of control systems, evaluation and monitoring, education and training and organizational safety performance at the individual and enterprise level. Like all models, its application will vary from one company or organization to another, given existing knowledge, skills, and attitudes. The model also has research potential as a way of organizing inquiry and investigation into the further interrelationships among these four key operational categories.

## 10. Conclusions

The introduction to this article noted the similarities in the epistemological principles behind finding Pluto for modern astronomers and finding safety performance incidents or characteristics in the universe of organizational OHS. The following are three key conclusions from this article:

1. The goal of predictive analytics needs to align with models of positive safety performance rather than models of failures in safety performance
2. The exploration of predictive analytics is essentially cultural, beginning with quantitative data shaped to reflect qualitative ends
3. The integration of predictive analytics looks differently against the backdrop of industrial organizational and operation units: each plays a part through extracting, communicating, managing, and capitalizing.

The principles of data mining for leading indicators, variables, clues, and intuitions of desired safety outcomes is a shared concern that motivates all professionals in modern, human-resource managed organizations. Like the bending of the solar system in response to a gravitational change, the predictive big-data analytics model presented here offers a way to address analytics in an integrated way. Developments in our understanding of how to integrate new perspectives on core safety operations can be a small step for predictive analytics and a giant leap for OHS professionals.

## References

[1] History.com Editors. Pluto discovered — HISTORY. In: HISTORY [Internet]. 3 Mar 2010 [cited 11 Feb 2019]. Available: https://www.history.com/this-day-in-history/pluto-discovered.
[2] Schultz G. Using advanced analytics to predict and prevent workplace injuries — occupational health & safety. [cited 10 Sep 2020]. Available: https://ohsonline.com/Articles/2012/07/01/Using-Advanced-Analytics-to-Predict-and-Prevent-Workplace-Injuries.aspx.
[3] Huang L, Wu C, Wang B, Ouyang Q. A new paradigm for accident investigation and analysis in the era of big data. Process Saf Prog 2018;37:42–8.
[4] Wang B, Wu C. Demystifying safety-related intelligence in safety management: some key questions answered from a theoretical perspective. Saf Sci 2019;120:932–40.
[5] Wang B, Wu C. Safety informatics as a new, promising and sustainable area of safety science in the information age. J Clean Prod 2020;252:119852.

[6] Ouyang Q, Wu C, Huang L. Methodologies, principles and prospects of applying big data in safety science research. Saf Sci 2018;101:60–71.
[7] Tripanagnostopoulos Y, Nousia T, Souliotis M, Yianoulis P. Hybrid photovoltaic/thermal solar systems. Solar Energy 2002;72:217–34.
[8] Dekker S. Safety differently: human factors for a new era. CRC Press; 2014.
[9] Hollnagel E. Safety management—looking back or looking forward. Resilience Eng Perspect 2016;1. Available: https://www.taylorfrancis.com/books/e/9781351903929/chapters/10.4324%2F9781315244396-11.
[10] Wick C. Knowledge management: communication of technical information, vol. 47. Technical Communication; 2000.
[11] Balfe N, Leva MC, Ciarapica-Alunni C, O'Mahoney S. Total project planning: integration of task analysis, safety analysis and optimisation techniques. Saf Sci 2017;100:216–24.
[12] Pabico C. Positive healthcare practice environments: implications for patient safety and outcomes. Health Care Curr Rev 2016;4. https://doi.org/10.4172/2375-4273.C1.017.
[13] Li J, Hale A. Identification of, and knowledge communication among core safety science journals. Saf Sci 2015;74:70–8.
[14] Rao TR, Mitra P, Bhatt R, Goswami A. The big data system, components, tools, and technologies: a survey. Knowl Inf Syst 2018. Available: https://link.springer.com/article/10.1007/s10115-018-1248-0; 2018.
[15] Li J, Hale A. Output distributions and topic maps of safety related journals. Saf Sci 2016;82:236–44.
[16] Huang L, Wu C, Wang B, Ouyang Q. Big-data-driven safety decision-making: a conceptual framework and its influencing factors. Saf Sci 2018;109:46–56.
[17] Laney D. 3D Data management: controlling data volume, velocity and variety. Contract no: research note. 2015; vol. 6; 2001.
[18] Pfeffer J, Jeffrey P, Thomas D., Dee II Professor of Organizational Behavior Graduate School of Business Jeffrey Pfeffer. The human equation: building profits by putting people first. Harvard Business Press; 1998.
[19] Zacharatos A, Barling J, Iverson RD. High-performance work systems and occupational safety. J Appl Psychol 2005;90:77–93.
[20] Li J, Goerlandt F, Li KW. Slip and fall incidents at work: a visual analytics analysis of the research domain. Int J Environ Res Publ Health 2019;16. https://doi.org/10.3390/ijerph16244972.
[21] Dadich A, Fulop L, Ditton M, Campbell S, Curry J, Eljiz K, et al. Finding brilliance using positive organizational scholarship in healthcare. J Health Organisat Manag 2015;29:750–77.
[22] Sanchez-Pi N, Martí L, Molina JM, Bicharra García AC. Information fusion for improving decision-making in big data applications. In: Pop F, Kołodziej J, Di Martino B, editors. Resource management for big data platforms: algorithms, modelling, and high-performance computing techniques. Cham: Springer International Publishing; 2016. p. 171–88.
[23] Camplin JC. Demonstrating safety performance through leading indicators. ASSE Professional Development Conference and. 2012. Available, https://www.onepetro.org/conference-paper/ASSE-12-601.
[24] Dyck D, Roithmayr T. Great Safety Performance: an improvement process using leading indicators. AAOHN J 2004;52:511–20.
[25] Kamaletdinova GR, Onevsky MP, Skvortsov SA. Human safety in a man-made ecosystem. Int J Safe 2018;8:406–12.
[26] Johnsen SO. Risks, safety and security in the ecosystem of smart cities. In: Svalova V, editor. Risk assessment. InTech; 2018.
[27] Crocker J. Safety in numbers: shifting from egosystem to ecosystem. Psychol Inq 2011;22:259–64.
[28] Hollnagel E, Wears RL, Braithwaite J. From safety-I to safety-II: a white paper. The resilient health care net: published simultaneously by the University of Southern Denmark. Australia: University of Florida, USA, and Macquarie University. 2015. Available: http://www.qpsolutions.vn/cgi-bin/Document/Safety%20II%20WhitePaper.pdf.
[29] Shi-Nash A, Hardoon DR. Data analytics and predictive analytics IN the era OF big data. In: Geng H, editor. Internet of things and data analytics handbook. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2017. p. 329–45.
[30] Wang B, Wu C, Huang L, Kang L. Using data-driven safety decision-making to realize smart safety management in the era of big data: a theoretical perspective on basic questions and their answers. J Clean Prod 2019;210:1595–604.
[31] Dasgupta T, Naskar A, Saha R, Dey L. Extraction and visualization of occupational health and safety related information from open web. In: 2018 IEEE/WIC/ACM international conference on web intelligence (WI). ieeexplore.ieee.org; 2018. p. 434–9.
[32] Badri A, Boudreau-Trudel B, Souissi AS. Occupational health and safety in the industry 4.0 era: a cause for major concern? Saf Sci 2018;109:403–11.
[33] Deloitte & Touche LLP. Workplace safety analytics. Available: https://www2.deloitte.com/content/dam/Deloitte/ca/Documents/Analytics/ca-en-analytics-workplace-safety-analytics.pdf.
[34] Li J, Reniers G, Cozzani V, Khan F. A bibliometric analysis of peer-reviewed publications on domino effects in the process industry. J Loss Prev Process Ind 2017;49:103–10.
[35] Hunter KO, Wolf EM. Cracking the code of process safety culture with organizational network analysis. Process Saf Prog 2016;35:276–85.
[36] Ancel E, Shih AT, Jones SM, Reveley MS, Luxhøj JT, Evans JK. Predictive safety analytics: inferring aviation accident shaping factors and causation. Null 2015;18:428–51.
[37] Lian Y, Zhang G, Lee J, Huang H. Review on big data applications in safety research of intelligent transportation systems and connected/automated vehicles. Accid Anal Prev 2020;146:105711.
[38] Corrigan S, Kay A, Ryan M, Brazil B, Ward ME. Human factors & safety culture: challenges & opportunities for the port environment. Saf Sci 2020;125:103854.
[39] Salas E, DiazGranados D, Klein C, Burke CS, Stagl KC, Goodwin GF, et al. Does team training improve team performance? A meta-analysis. Hum Factor. 2008;50:903–33.
[40] Global occupational safety and health. 13 Nov 2018 [cited 10 Sep 2020]. Available: https://www.cdc.gov/niosh/topics/global/default.html.
[41] Jørgensen K. A systematic use of information from accidents as a basis of prevention activities. Saf Sci 2008;46:164–75.
[42] Government of Canada, Canadian Centre for occupational health, safety. Workplace health and well-being promotion – Getting started. [cited 12 Feb 2019]. Available: http://www.ccohs.ca/.
[43] Stanton JM, Balzer WK, Smith PC, Parra LF, Ironson G. A general measure of work stress: the stress in general scale. Educ Psychol Meas 2001;61:866–88.
[44] Dyck DEG. Occupational health & safety: theory, strategy & industry practice. Canada: LexisNexis; 2015.
[45] Goraya A, Amyotte PR, Khan FI. An inherent safety–based incident investigation methodology. Process Saf Prog 2004;23:197–205.
[46] Donaldson C. Predicting safety: analytics comes to OHS. OHS Professional. 2017. Available: http://safetyontap.com/wp-content/uploads/2017/07/SIA-OHS-Professional-March-2017-analytics-comes-to-OHS.pdf.
[47] Trevisan DG, Sanchez-Pi N, Marti L, Garcia ACB. Big data visualization for occupational health and security problem in oil and gas industry. Human interface and the management of information information and knowledge design. Springer International Publishing; 2015. p. 46–54.
[48] Provan DJ, Dekker SWA, Rae AJ. Bureaucracy, influence and beliefs: a literature review of the factors shaping the role of a safety professional. Saf Sci 2017;98:98–112.
[49] Horton J, Cameron A, Devaraj D, Hanson RT, Hajkowicz SA. Workplace Safety Futures: the impact of emerging technologies and platforms on work health and safety and workers' compensation over the next 20 years. Canberra: CSIRO. 2018. Available: https://www.inxsoftware.com/media/transfer/doc/workplacesafety.pdf.
[50] Gwilt I, Rolph J, Eimontaite I, Cameron D, Aitken J, Mokaram S, et al. Cobotics: developing a visual language for human-robotic collaborations. In: Proceedings of the cumulus conference 2018. Available: https://www.researchgate.net/profile/Jonathan_Aitken2/publication/324079462_Cobotics_developing_a_visual_language_for_human-robotic_collaborations/links/5af2db35a6fdcc0c0304fdc6/Cobotics-developing-a-visual-language-for-human-robotic-collaborations.pdf; 2018.
[51] Li X, Yi W, Chi H-L, Wang X, Chan APC. A critical review of virtual and augmented reality (VR/AR) applications in construction safety. Autom Constr 2018;86:150–62.
[52] Chan Albert PC, Javed Arshad Ali, Wong Francis KW, Hon Carol KH. Improving safety communication of ethnic minorities in the construction industry of Hong Kong. ICCREM 2014. https://doi.org/10.1061/9780784413777.056.
[53] Guo SY, Ding LY, Luo HB, Jiang XY. A Big-Data-based platform of workers' behavior: observations from the field. Accid Anal Prev 2016;93:299–309.
[54] Branson N. Re-weighting the OHS and LFS national household survey data to create a consistent series over time: a cross entropy estimation approach. Southern Africa Labour and Development Research Unit, University of Cape Town; 2009.
[55] Bifet A. Mining big data in real time. Informatica 2013;37. Available: http://www.informatica.si/index.php/informatica/article/download/428/432.
[56] Dibenedetto DV. Identifying requirements for an occupational health and safety information system. AAOHN J 2004;52:368–72.
[57] Salazar MK, Kemerer S, Amann MC, Fabrey LJ. Defining the roles and functions of occupational and environmental health nurses. Results of a national job analysis. AAOHN J 2002;50:16–25.
[58] Franchini L, Caillaud E, Nguyen P, Lacoste G. Workload control of human resources to improve production management. Int J Prod Res 2001;39:1385–403.
[59] Thissen-Roe A. Adaptive selection of personality items to inform a neural network predicting job performance. University of Washington. 2005. Available: http://iacat.org/sites/default/files/biblio/th05-01.pdf.
[60] Strohmeier S, Piazza F. Domain driven data mining in human resource management: a review of current research. Expert Syst Appl 2013;40:2410–20.
[61] Morabito V. Big data and analytics innovation practices. In: Morabito V, editor. Big data and analytics: strategic and organizational impacts. Cham: Springer International Publishing; 2015. p. 157–76.
[62] Johnson J, Haegeli P, Hendrikx J, Savage S. Accident causes and organizational culture among avalanche professionals. J Outdoor Recreation Tourism 2016;13:49–56.
[63] Bail CA. The cultural environment: measuring culture with big data. Theory Soc 2014;43:465–82.
[64] Pence J, Mohaghegh Z, Kee E, Yilmaz F, Grantom R, Johnson D. Toward monitoring organizational safety indicators by integrating probabilistic risk assessment, socio-technical systems theory, and big data analytics. In: Proceedings of 12th international topical meeting on probabilistic safety assessment and analysis (PSAM12). psam12.org 2014. Available: http://psam12.org/proceedings/paper/paper_549_1.pdf; 2014.
[65] Grossman R, Siegel K. Organizational models for big data and analytics; 2014. Available: https://papers.ssrn.com/abstract=2458909.
[66] Davenport TH. Competing on analytics. Harv Bus Rev 2006;84. 98–107, 134.

[67] Fernández FB, Pérez MÁS. Analysis and modeling of new and emerging occupational risks in the context of advanced manufacturing processes. Procedia Eng 2015;100:1150–9.
[68] Felt M. Social media and the social sciences: how researchers employ Big Data analytics. Big Data Soc 2016;3. 2053951716645828.
[69] Segall RS, Niu G. Overview of big data and its visualization. In: Global IGI, editor. Handbook of research on big data storage and visualization techniques. IGI Global; 2018. p. 1–32.
[70] McCue C. Data mining and predictive analytics in public safety and security. IT Prof 2006;8:12–8.
[71] Robson LS, Clarke JA, Cullen K, Bielecky A, Severin C, Bigelow PL, et al. The effectiveness of occupational health and safety management system interventions: a systematic review. Saf Sci 2007;45:329–53.
[72] Robson LS, Amick III BC, Moser C, Pagell M, Mansfield E, Shannon HS, et al. Important factors in common among organizations making large improvement in OHS performance: results of an exploratory multiple case study. Saf Sci 2016;86:211–27.
[73] Cooren F, Kuhn T, Cornelissen JP, Clark T. Communication, organizing and organization: an overview and introduction to the special issue. Organ Stud 2011;32:1149–70.
[74] Wilhoit ED. Space, place, and the communicative constitution of organizations: a constitutive model of organizational space. Commun Theory 2018;28:311–31.
[75] Badri A, Gbodossou A, Nadeau S. Occupational health and safety risks: towards the integration into project management. Saf Sci 2012;50:190–8.