



Collection and Interpretation of Big Data

YPSW AIHA Annual Meeting

Total Worker Exposure
January 25th, 2019

© Copyright 2018 by ERM Worldwide Group Limited and/or its affiliates ('ERM'). All Rights Reserved.
No part of this work may be reproduced or transmitted in any form or by any means, without prior written permission of ERM.

The business of sustainability



Data is the new asset

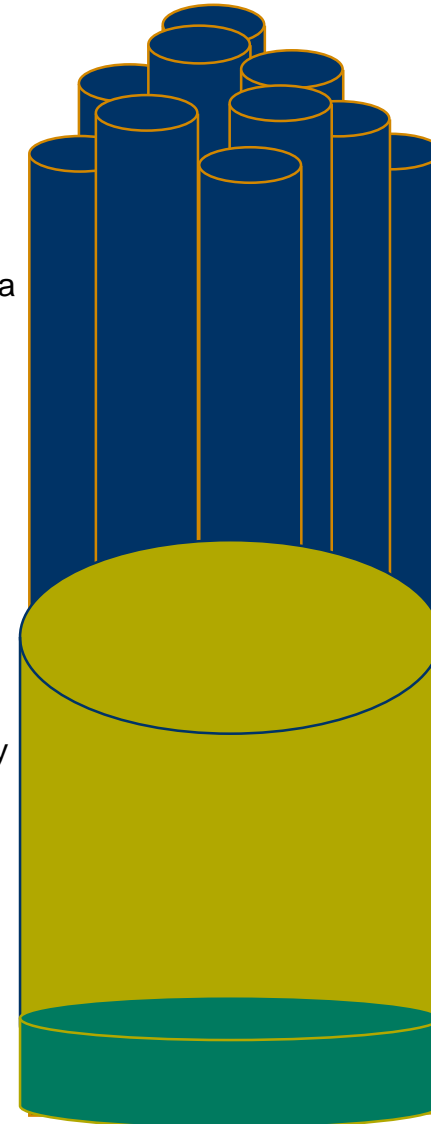


The world's most valuable resource is no longer oil, but data

Variety – new type of data constantly being utilized

Velocity – continuously changing data

Volume – static, data points, records



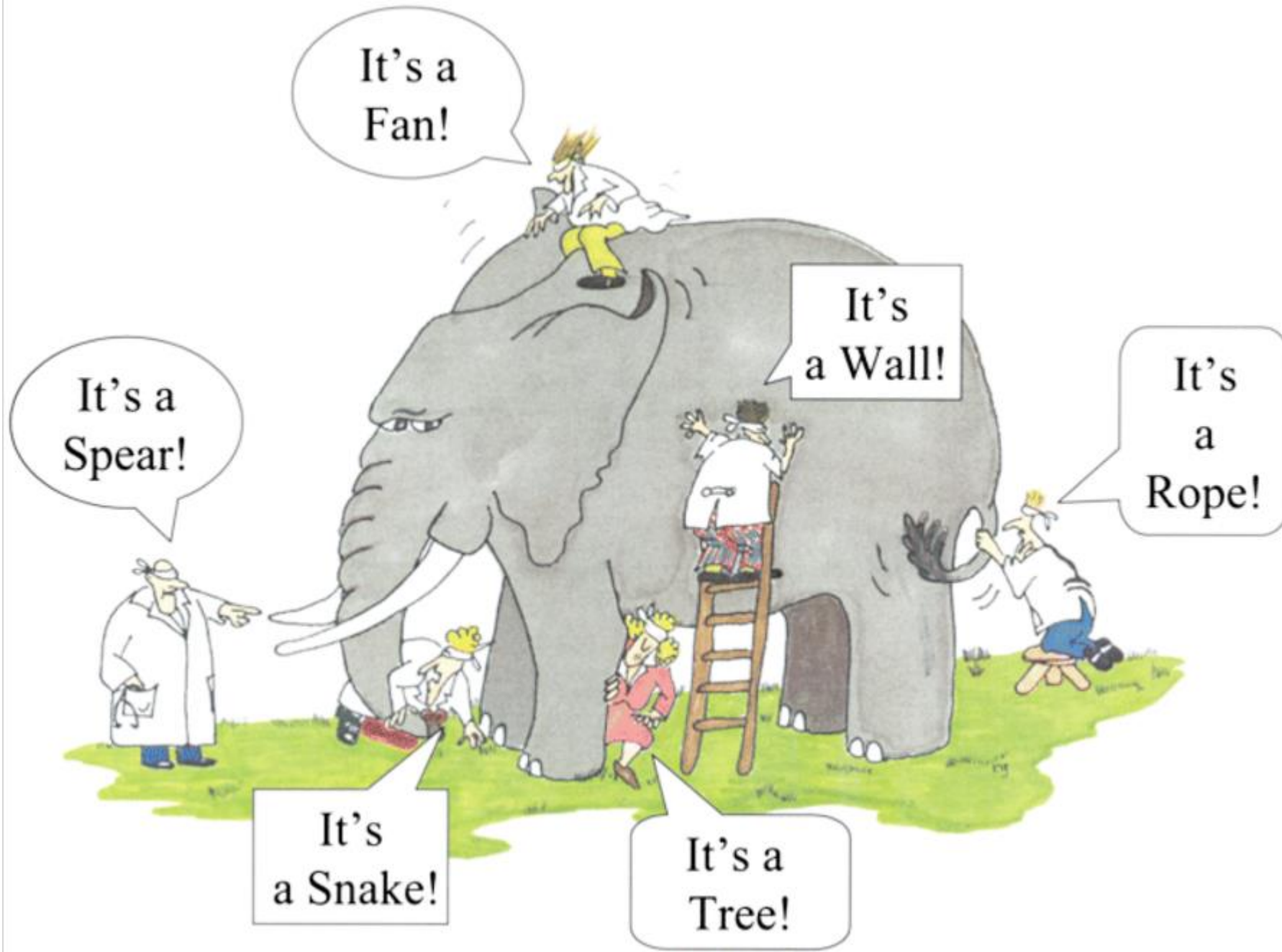
Geospatial, Imagery, Voice, Video, Text and more forms of new data which can be processed at low cost now

IOT, Wearables, AR, VR and continuously changing data sets are being leveraged by Production plants

Traditional Data sets which reside in Databases and Excel sheets

Agenda

- Definitions ... what do these words mean
- Industrial Hygiene (IH) Data Value Chain ... how does this impact IH
- The Power (and Fury) of Analytics ... what do you need to watch out for
- Toolkits and examples ... what can we do about it



Key Definitions

■ Big Data

- Big data is a term that describes the large volume of data – both structured and unstructured – that inundates a business on a day-to-day basis. But it's not just the amount of data that's important. It is the three Vs – **Volume; Velocity and Variety** as well as **Variability and Complexity** that define the term

■ Machine Learning

- A method of data analysis that automates analytical model building. It is a branch of artificial intelligence based on the idea that systems can learn from data, identify patterns and make decisions with minimal human intervention.

■ Predictive Analytics

- Encompasses a variety of statistical techniques from data mining, predictive modelling, and machine learning, that analyze current and historical facts to make predictions about future or otherwise unknown events.

The Role of Data

The 4th Industrial Revolution Started in 2013 (Industry 4.0) – Cyber Physical Systems and takes over from the 3rd Revolution on Computer and Automation

Data waster

Collects data but severely underuse them



Data collector

Collects data but do not consistently maximize their value



Aspiring data manager

Understands value of data and marshals resources to take better advantage



Strategic data manager

Has well-defined data-management strategies that focus on collecting and analyzing the most valuable data



Digital Maturity Model



Digital Maturity Level 1 Moving from Analog

L1 Characteristics

- Isolated Data Sets
- Pen & Paper → Excel
- Bespoke Reports
- Reports On Hard Copies/Hard Drives
- Lack Of Transparency

**INCONSISTENT AND
INEFFICIENT?**



Digital Maturity Level 2 Making the Digital Leap

L2 Characteristics

- Digital Data Collection
- Single Source Of Truth
- Speed Up Decision Making Process (Real Time)
- Consistent And Efficient Reporting

**RELIABLE AND
EFFICIENT?**



Digital Maturity Level 3 Integrating on Digital

L3 Characteristics

- Connecting Business Data Sources
- Data Insights Leading To Optimization
- Better Visualization (3D) → Stakeholders
- Efficient Decision Making Process

**INTEGRATED AND
INSIGHTFUL?**



Digital Maturity Level 4 Into Data Mining

L4 Characteristics

- Predictive Analytics
- Machine Learning For Data Analysis
- Automated Decision Making

To Artificial Intelligence,
Virtual Reality And Beyond

**FORWARD-LOOKING AND
PROGRESSIVE?**

Data analytics overview

Descriptive analytics

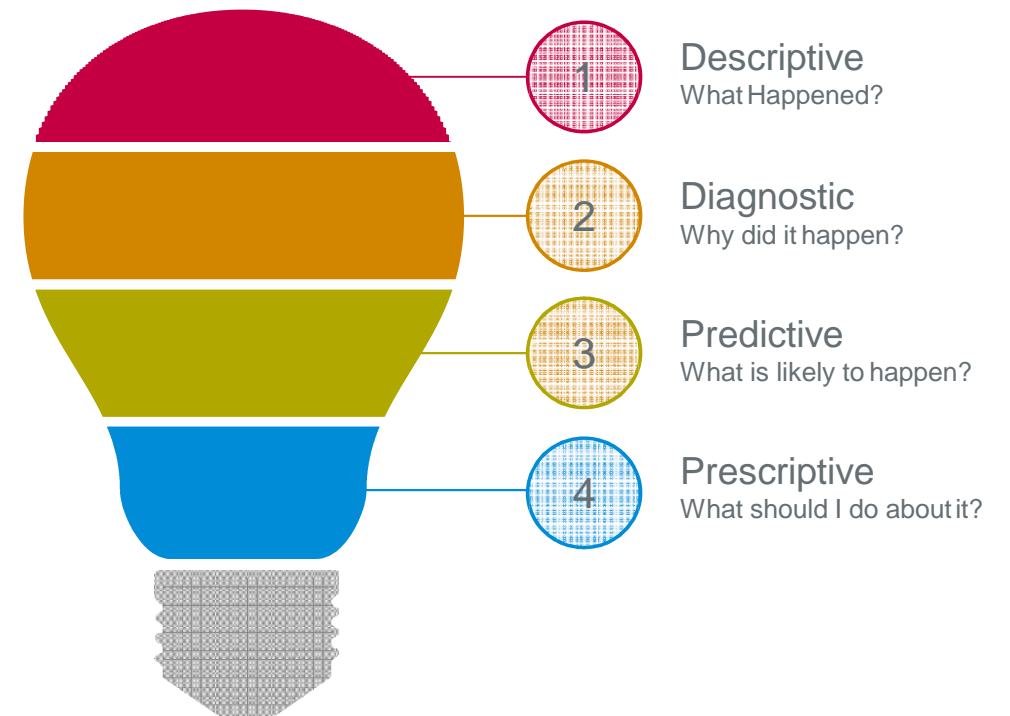
- Describes or summarizes raw data and makes it interpretable

Predictive analytics

- Helps to understand the future and provide actionable insights

Prescriptive analytics

- Quantifies the effect of future decisions



Industrial Hygiene Data Value Chain

Data Aggregation



- Equipment and Components to monitor
- Observation Paths
- Training
- Sampling
- Competency Tracking
- Field Devices
- Monitoring

Data Integration / Analysis



- Data Curation
- Integration to other data sets
- Key findings to help drive actions for the other functions

Data Utilization



- Analyze to identify trends
- Define key actions
- Regulatory and In-House Reports

Data Aggregation - KISS

- **Collection and Annotation needs to be the current focus of attention – keeping it simple will be the key**
 - Utilize sensors currently available
 - Leverage what operations is already putting out in place
- **However a successful and SMART IH program needs to focus on the preparation and organization**
 - Clear definition of equipment and components to monitor
 - Training programs – resource identification and planning
 - Right to Left Thinking – define the “end game”
- **Prepare and Organize to ensure that Collect and Annotate is successful**

Data Integration / Curation

Avoid data silos

- Standalone systems add to operating complexity
- Unable to fully leverage the value of the data
- Integration can be complex – but ways to manage complexity



Cleansing / Curating the data for further analysis:

- Data Profiling
- Data Quality Safeguards
- Data Cleansing
- Data Quality Standards and Training

Recognizing “Dirty Data”

- **Incorrect Data** – value is not valid
- **Inaccurate Data** – value is not true (still may be valid)
- **Inconsistent Data** – same value appears in two places (with minor differences)
- **Incomplete Data** – the data set does not have all the information required for the task or analysis
- **Business Rule Violations** – value is valid and accurate but violates a business rule
- **Nonintegrated Data** - data is not connected appropriately with other data sets which makes analysis and comparisons difficult

Art and Science of Data Curation

Data Profiling

- Uncover your data defects with some data archeology – analyze the data for correctness, completeness, uniqueness, consistency and reasonability

Data Cleansing

- After profiling is complete focus on cleaning the critical and important areas – need to prioritize effort

Data Quality Safeguards

- Establish logic, programming, checks to prevent future “dirty data”

Data Quality Standards and Training

- Establish governance process, data stewards and develop training to drive these standards and guidelines across the enterprise

Data Insights

Companies acquire complex and significant amounts of data over the operational lifetime of a site. ERM experience suggests data is rarely stored in a centralised, organised place, can be difficult to locate, is inconsistent, its purpose unclear and the true value of the data in managing risks lost.

ERM's approach is centred on ERM's Digital Foundation a, set of platforms and tools that with our subject matter experts allow us to:



Acquire, transform, and persist data



Apply data science techniques to find correlations within data

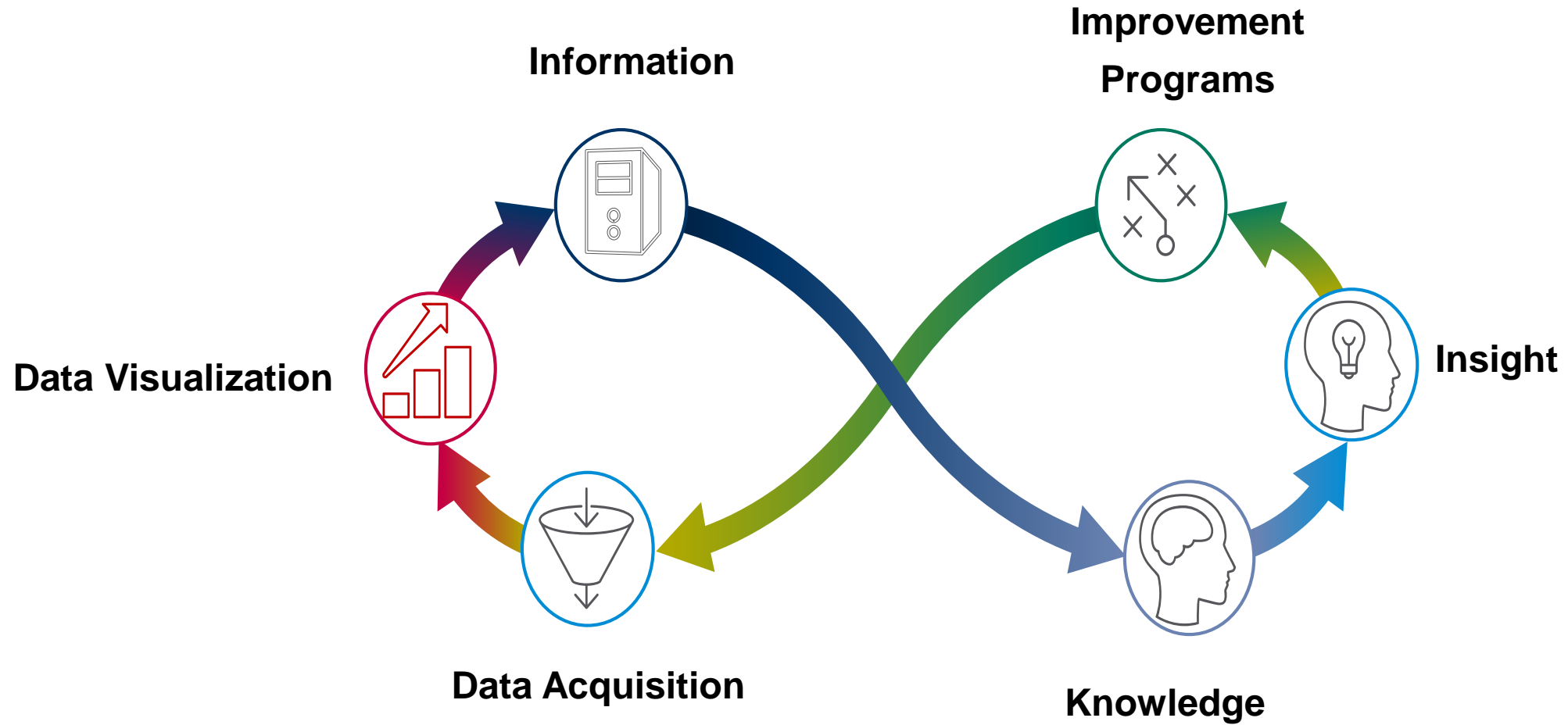


Apply consultative insight to determine the “so what” in context of the problem that you the client are trying to solve and deliver value

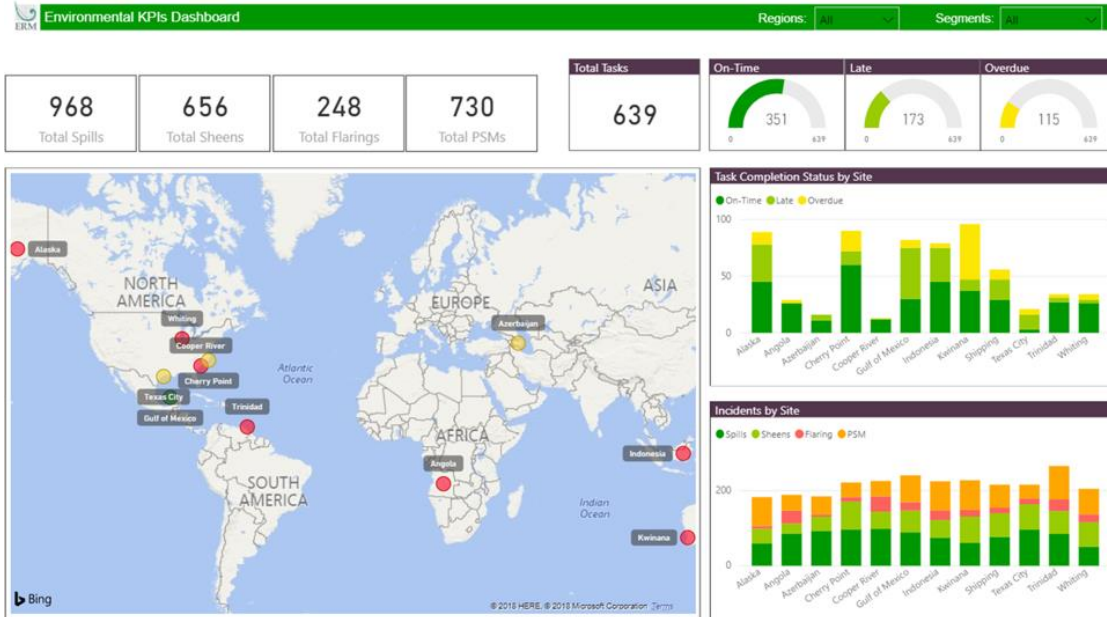


Establish foundation for move to mobile

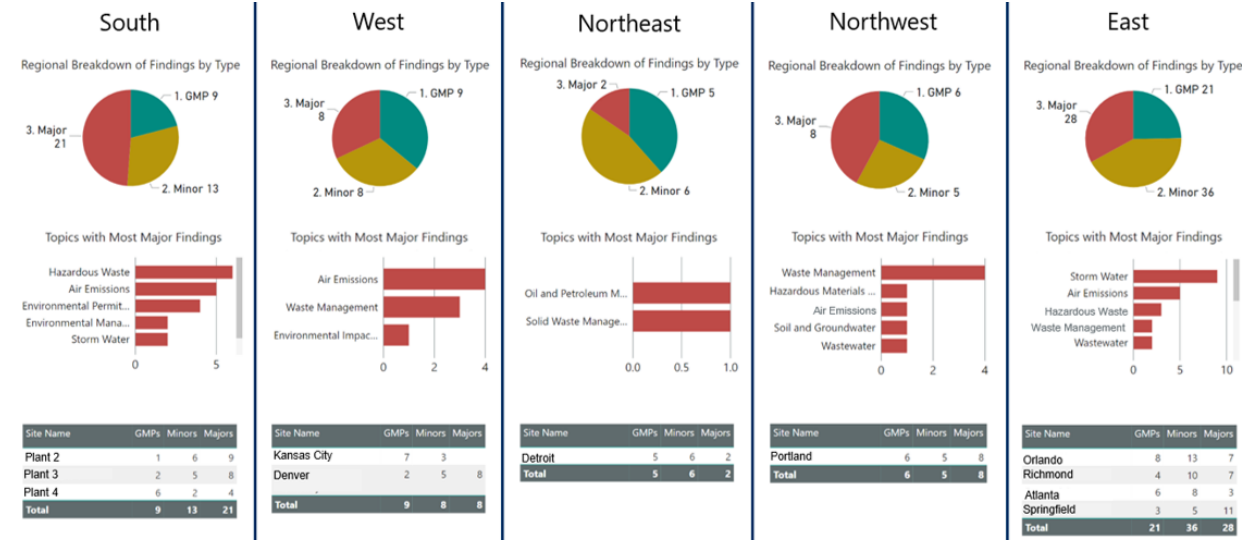
Virtuous cycle of analytics



Data Utilization



Audit Samples Dashboard



Performance Monitoring Dashboard

Power (and Fury) of Analytics

Prediction is very difficult ...

Especially if its' about the future. – Neils Bohr

- *Text message to chemical plant manager:* Chlorine leak expected on line 2 tomorrow. Inspect and repair.
- *High priority email and automatic call to coal mine superintendent:* 83% chance of roof fall on section 4. Evacuate immediately and take corrective actions.
- *Monthly notice to OSHA regional administrator:* HIGH PRIORITY INSPECTION ROSTER: Firms listed below have a greater than 80% probability of violations reflecting hazardous conditions requiring mitigation.

Barriers to implementation:

- Knowledge – don't have the data
- Skills – don't have the skills / technology to mine the data
- Motivation - not keen on acting on the insights
- Inertia - fine with current state so not sure the additional effort is worth it
- Privacy - GDPR and PII data

Source: Can Predictive Analytics help reduce workplace risk, Dr. Wagner, CDC Blog

Challenges to manage with big data analytics

- **Causation and Correlation - *Post hoc ergo propter hoc***
 - Data Scientists get focused on the statistical analysis of correlation without thoughtful reflection of causation
- **Sample Size and Data Collection**
 - Need to have a proper and representative sample
 - “Stuck in the data forest and lose sight of the smaller but more important trees”
 - Issue is starting to become “too much data” rather than “not enough data”
- **In-built biases in our algorithms**
 - Algorithms are written by human beings – biases do creep in
- **Privacy concerns**

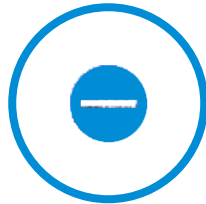
Four Pillars of Analytics Excellence

Confidence



Broad Datasets

Privacy



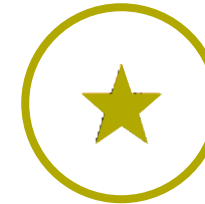
Guaranteed

Security



Certified Protected

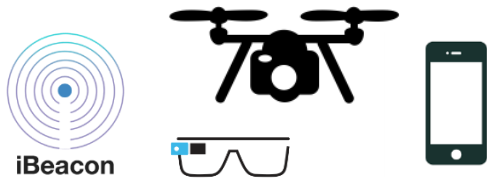
Expertise



Domain Knowledge

Key digital themes

Measuring & Monitoring



Wearables

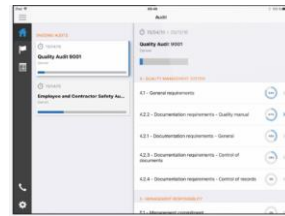
Augmented / Virtual Realities

Internet of Things & Sensors

Robotics / Drones

Mobility

Systems & Automation



Information Systems

Public Information

Data Lakes / Warehouse

Cognitive Intelligence

Blockchain

Reporting, Analytics & Communication



Analytics / Data Science

Visualization Tools

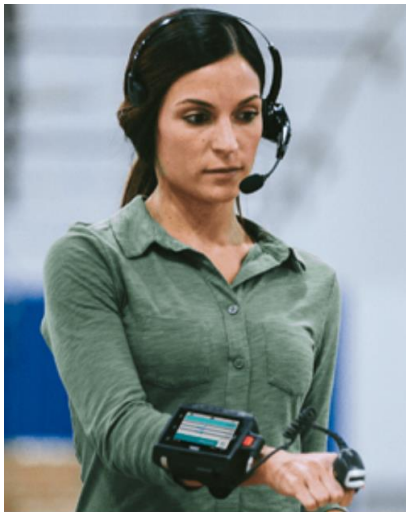
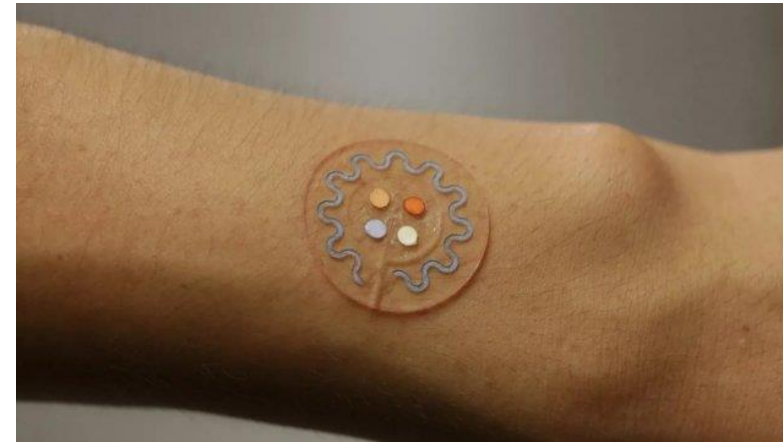
Social Media

Quick Look: Wearables

Avoiding Dehydration

To avoid dehydration, especially in the Middle East during Ramadan, BHGE is leveraging patches to measure sweat to monitor potential for dehydration. Expanding now to monitor fatigue and stress.

This patch is developed through a consortium of GE, USAF, and universities. A similar patch being developed by a consortium, including Gatorade



Understanding ergonomics

By 2018, 2MM employees will be required to wear health & fitness tracking devices as a condition of employment (per Gartner). Insurance companies, such as Liberty Mutual, are partnering with industry to identify risk areas, such as body movements of workers. Wearables capture movements (e.g., bending, reaching), which are analyzed and then the appropriate equipment and/or work processes are ordered

Quick Look: New Realities

VR Training for Improving Driver Safety

With large exposure of workers on the road, driving as part of their normal job activities, companies, such as UPS and Linde are leveraging VR training to allow employees to become more aware to hazards and risks. This same approach is leveraged highly in the Formula 1 space for impressive results. Research are showing that learnings / time-to-mastery are accelerated by two-thirds.



Remote Inspections – BASF

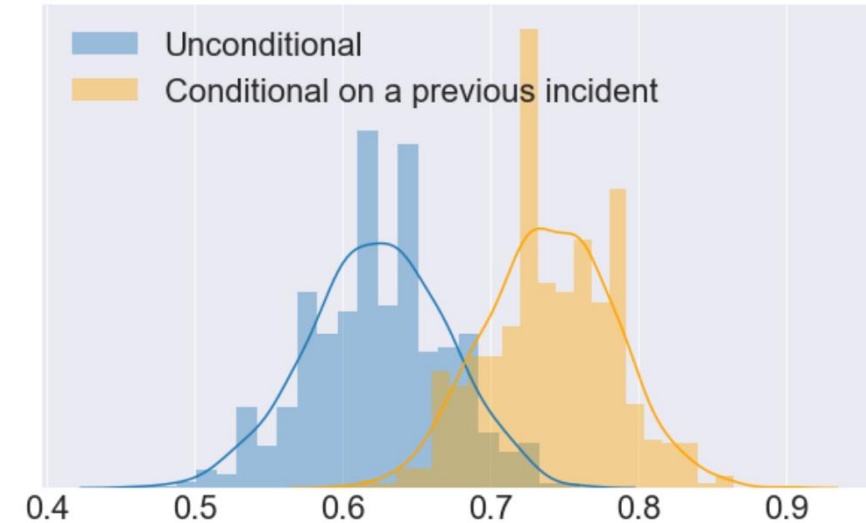
Inspecting and maintaining fire trucks in Germany was expensive - travel expenses, a week of downtime and additional costs. By leveraging AR/VR, the inspections are significantly improved by allowing technician to analyze the truck remotely and identify any parts prior to coming on-site to service the truck

Quick Look: Analytics

Generating Insights

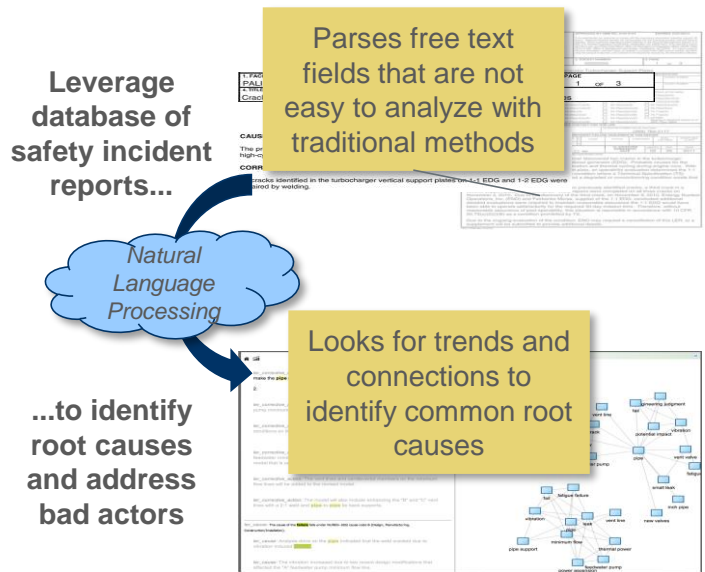
Reviewing existing incident data in combination with a number of different data elements (e.g., weather) and applying data science and subject matter expertise to understand the correlations

Probability of seeing at least one Actual Incident (serious or above) over the next 30 days

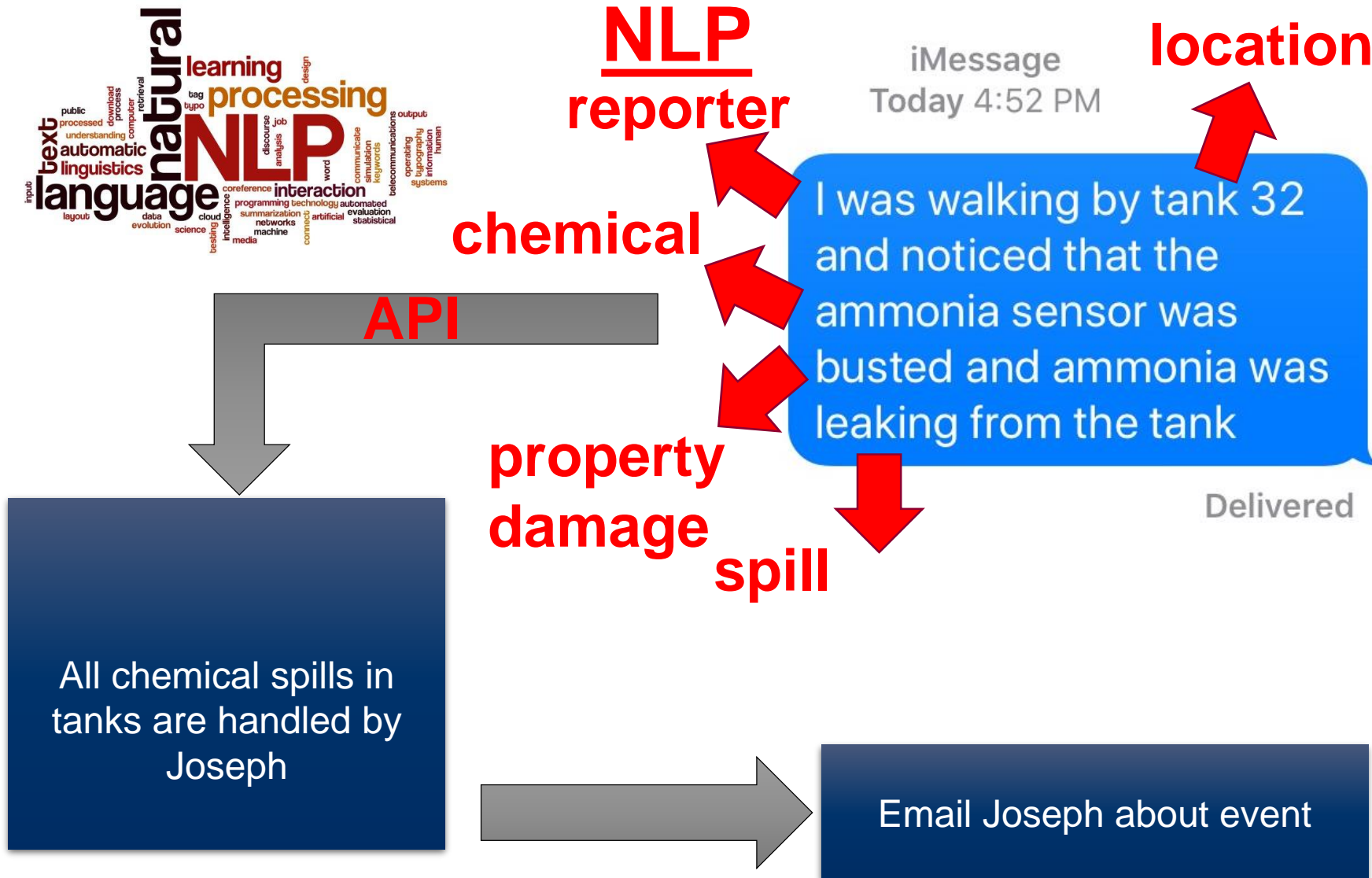


Natural Language Processing

Use of machine learning and natural language processing to improve the ability to analyze large volumes of data and identify hazards, which may be embedded in unstructured free text reports



Plain Speak to Useful Data?



Summary

- **Understand what data is available – KISS**
- **Focus on “Clean Data”**
- **Define the questions you are trying to answer – what is the end-point**
- **Leverage technology to improve your data collection & analysis**
- **Be willing to try – “fail fast” – be AGILE**
- **Understand how to scale**