



**CLEAR** LEARNING

CLEAR Call Webinar Series

# How regulatory agencies are using big data to predict non-compliance

Mark Parker  
Rick Fiene

[www.clearhq.org](http://www.clearhq.org)

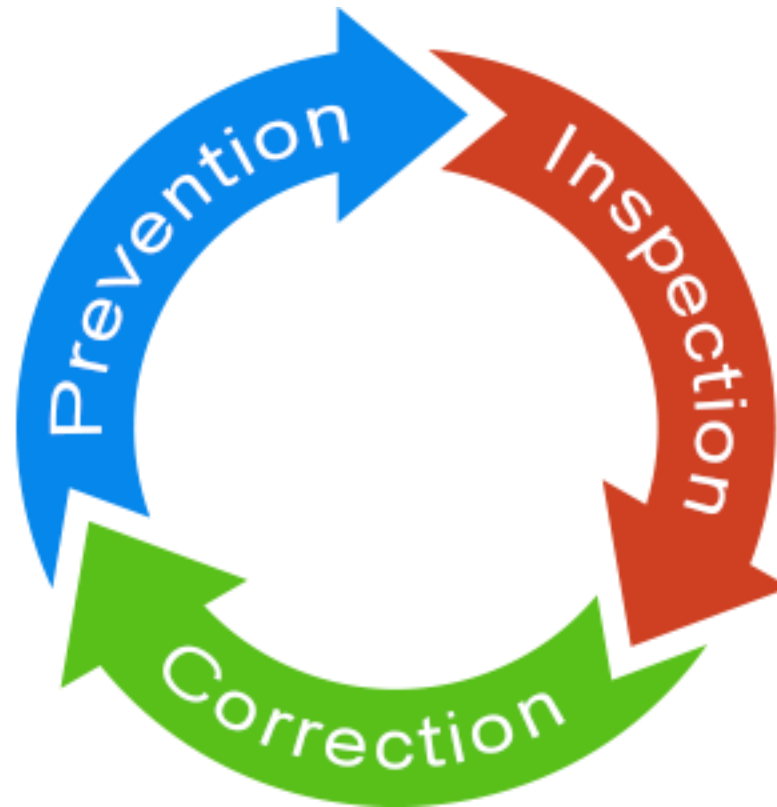
Council on Licensure, Enforcement and Regulation

# Background

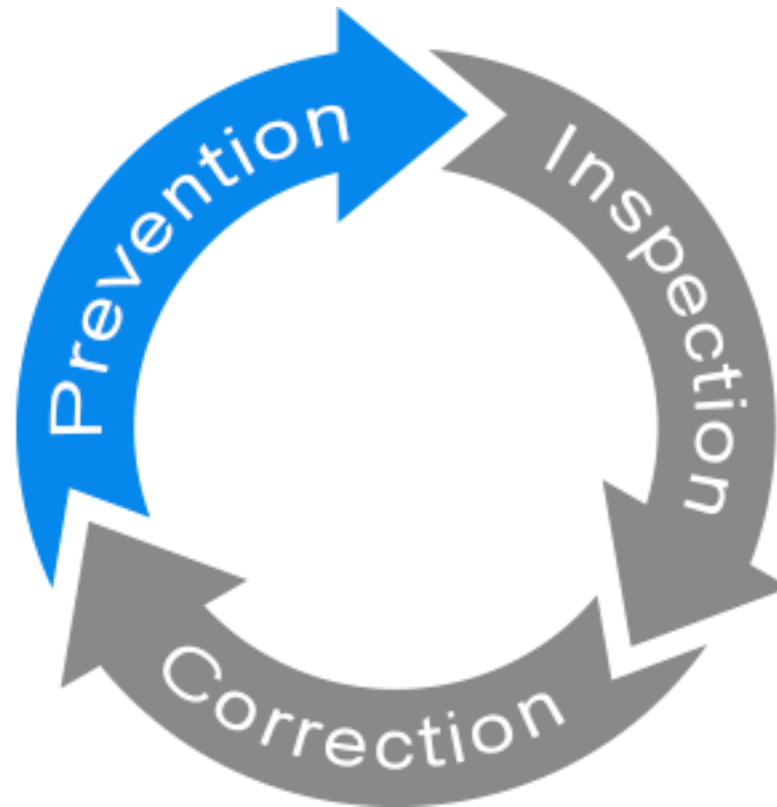


- ▶ Regulatory Science.
- ▶ How data systems have changed in past 50 years.
- ▶ Predictive Analytics.
- ▶ Focus more on the limitations of regulatory compliance data.
- ▶ Ceiling effect.
- ▶ Skewed data.
- ▶ Nominal.
- ▶ Little variance.

# Compliance Monitoring Cycle



# Compliance Monitoring Cycle



# What is Data Analytics

---



**Data analytics** is the practice of examining data to answer questions, identify trends and extract insights.



---

# Data Visualization



---

# Polling Question #1







The Leading Indicator of Non-compliance

# What is Big Data?



## The six Vs of big data

Big data is a collection of data from various sources, often characterized by what's become known as the 3Vs: *volume*, *variety* and *velocity*. Over time, other Vs have been added to descriptions of big data:

VOLUME	VARIETY	VELOCITY	VERACITY	VALUE	VARIABILITY
The amount of data from myriad sources.	The types of data: structured, semi-structured, unstructured.	The speed at which big data is generated.	The degree to which big data can be trusted.	The business value of the data collected.	The ways in which the big data can be used and formatted.
					



# Big Data Uses



- Social Media
  - Targeted Advertising
- Manufacturing
  - Minimizing Downtime
- National Security / Public Safety
  - Predicting Potential Threats and Crimes

# Regulatory Agency Focus



- Volume
  - Amount of Data
- Variety
  - Types of Data
- Veracity
  - Trustworthiness of Data



- Storage (and Disaster Recovery)
  - On-prem, Cloud-based and Hybrid
- Many Sources and Repositories
  - Databases, Email Servers, Shared Drives, etc.
- Accessibility
  - Data Visualization

# Variety



- Unstructured
  - Email, Word Documents, Fillable PDFs, etc.
- Semi-structured
  - Excel Spreadsheets
- Structured
  - Relational Databases



---

# Polling Question #2

Technologies used to Capture Inspection Results



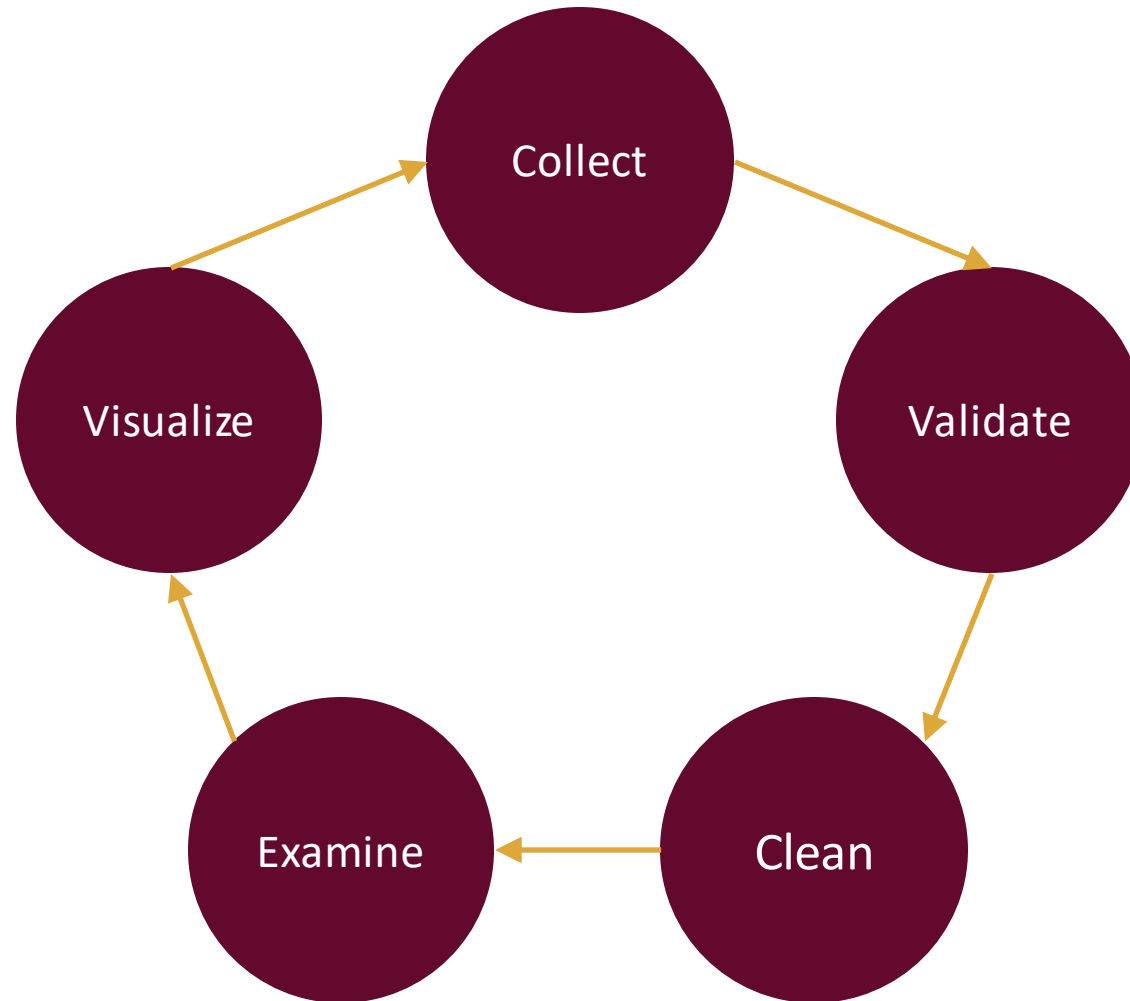
- **High** (Accurate, Consistent, Relevant and Unbiased)
  - Internet of Things (IoT) Input
    - Sensors, Cameras, Microphones, etc.
  - Picklist Input
    - Dropdown Lists, Radio Buttons, etc.
  - Freeform Input
    - Typing, Voice-to-Text, etc.
- **LOW** (Inaccurate, Inconsistent, Irrelevant and/or Biased)



---

# Data Entry Example

# Data Analysis Process





# Data Analytics in Compliance



- **Descriptive**
  - What happened?
- **Diagnostic**
  - Why did it happen?
- **Predictive**
  - What might happen in the future?
- **Prescriptive**
  - What should we do next?

# Methods and Techniques



- Time Series Analysis
  - Month-to-Month, Quarter-to-Quarter, Year-to-Year
- Cohort Analysis
  - Region, Supervisor, Inspector
- Regression Analysis
  - Duration of License, Years of Experience, Staff Turnover



---

# Data Visualization Example



---

# Polling Question #3

Management Training for Licensees

# Key Takeaways



- Takes time but doesn't have to be expensive
- Data visualization is necessary for exploring large datasets
- Accurate and consistent data collection is what makes predicting non-compliance possible

## Contact Information

### **Dr Richard Fiene**

RIKI LLC – Research Institute  
for Key Indicators

717-598-8908

Rfiene@RIKInstitute.com

### **Mark Parker**

Outlier Technologies

513-792-8072

mark.parker@outliertech.com

[www.linkedin.com/in/markparker-outliertech](http://www.linkedin.com/in/markparker-outliertech)

## **CLEAR**

108 Wind Haven Dr., Ste. A | Nicholasville, KY 40356 | (859) 269-1289 | [www.clearhq.org](http://www.clearhq.org)