AI for Health & Safety

Practical guidance for applying data science techniques in health & safety

Scott Kimbleton | **IBM**

Global AI Solutions Deployment Leader
Energy & Natural Resources

Graziella Caputo | **IBM**

Sr. Data Scientist & Manager
Data Science Elite

Hazards31 IChemE Safety Centre **IChemE** ADVANCING CHEMICAL ENGINEERING WORLDWIDE

November 18, 2021

Contents

- Why AI for HSE now?

- Taking a systemic view to apply data science at scale

- Practical aspects of data science for health & safety

- Summary, where to learn more, let's connect!

IBM

POLL!

Please identify <u>in the chat </u>which role your activities most closely align.

The four (4) main roles for today's topic:

- End User
- Subject Matter Expert
- Data Scientist
- Digital Transformation Leader

Other roles certainly exist in a comprehensive transformation programs but we've selected these four roles most relevant for today's topic

IBM

What is the value of AI for health & safety?

Pain Points

- Stagnating or declining operational excellence (OE) performance metrics

- Limited sharing of incident information across sites

- Intensive manual effort to gather and research incident information, for each hazard study

- Poor information for basic analysis, let alone predictive modeling; no objective measures of quality

AI Benefits

- *Uncover hidden trends* in behaviors and *undetected hazards*

- Automate information collection and quality improvement *at the point of capture*

- *Automatically find and visualize* key items such as hazard conditions, activities, incident causes and controls *found in free text*

- More accurately *identify situations with a higher than normal risk profile*

IBM

Imagine if....

Bob suffered a serious cut to his hand, the usual incident response happened.
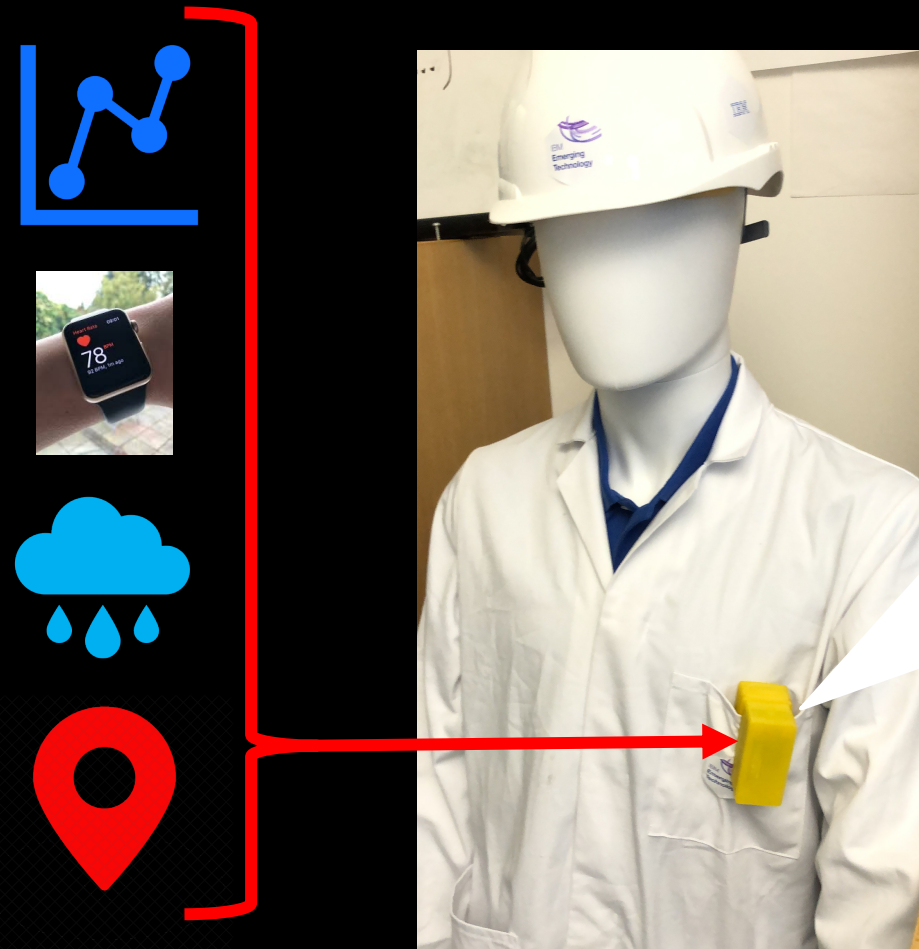
But, just before Mary started work:
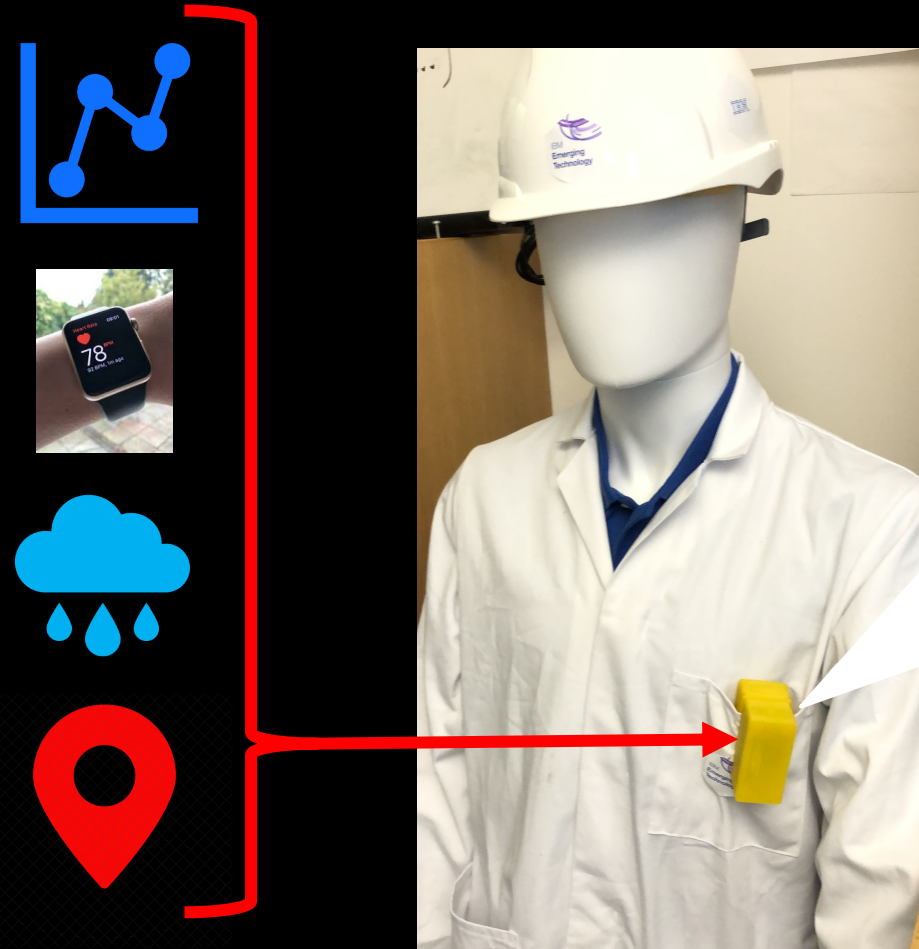
Hi Mary, its late in the day and I see you're about to do a manual handling task, did you know that someone cut their hand last week in a similar task, please wear cut proof gloves.

Imagine if....

Bob suffered a serious cut to his hand, the usual incident response happened.

But, just before Mary started work:

Hi Mary, its late in the day and I see you're about to do manual handling task, did you know that someone cut their hand last week in a similar task, please wear cut proof gloves.

1 – Risk factor (raise/lower total risk)
2 – High risk task (base risk)
3 – Prior or similar events
4 – Mitigating actions

For AI success, organizations must navigate the paradox of the worker safety *Information Blackhole* and the *Data Deluge*

Information Blackhole

- Entered into systems because we have to

- Rarely feeds back into the real world

- Not shared across organisations

- Is hard to find, especially out of the office

Data Deluge

- Decision fatigue - more than 200 conscious decisions a day...just on what to eat!

- Volume of data generation is outpacing infrastructure and tools to support

- More and more data sources; little integration

IBM

Why AI for Health & Safety now?

Algorithmic advances

Data proliferation

Increases in computing power & storage

1958 – Frank Rosenblatt develops the first self-learning algorithm

1989 – Birth of CNNs for image recognition; paper describes how CNNs are well suited for shape-recognition tasks.

2004 – Youtube debuts; 100M views a day within 18 months

2005 – iPhone launched; around-the-clock consumption and creation of data & content

1997 - Increase in computing power drives IBM's Deep Blue victory over Garry Kasparov

2005 – Cost of of disk storage drops 99.7% in 10 years (1GB, $277 to $0.79)

2011 - IBM's Watson beats *Jeopardy!*

2012 - Deep-learning system wins renowned image-classification contest for the first time

2017 - Electronic-device users generate 2.5 quintillion bytes of data per day

IBM

Advanced technology alone doesn't guarantee success – the right teams executing value *generating* and value *enabling* activities is key.

Value *Generating* - Addressing a specific business scenario

Shanthi
Process Engineer, **End User**

Alexandre
Supply Chain

Marcio
Operations & Maintenance

**Philip**
Data Scientist

Susan
Quality Specialist

Mark
Delivery Coordinator

Subject Matter Experts

IBM

Advanced technology alone doesn't guarantee success – the right teams executing value *generating* and value *enabling* activities is key.

Value *Generating* - Addressing a specific business scenario

Value *Enabling* - Replicating success across *100's of scenarios*

Shanthi
Process Engineer, **End User**

Alexandre
Supply Chain

Marcio
Operations & Maintenance

Philip
Data Scientist

Kai
Transformation Leader

Susan
Quality Specialist

Mark
Delivery Coordinator

Subject Matter Experts

IBM
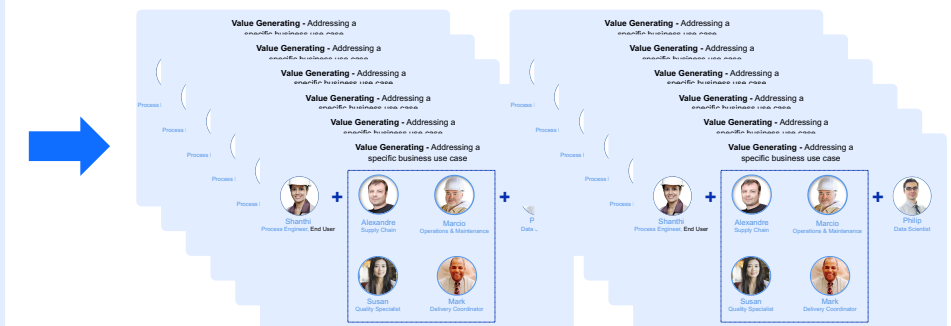
Improving your odds of achieving data science success shouldn't be a guessing game. There are four key frameworks to guide you.
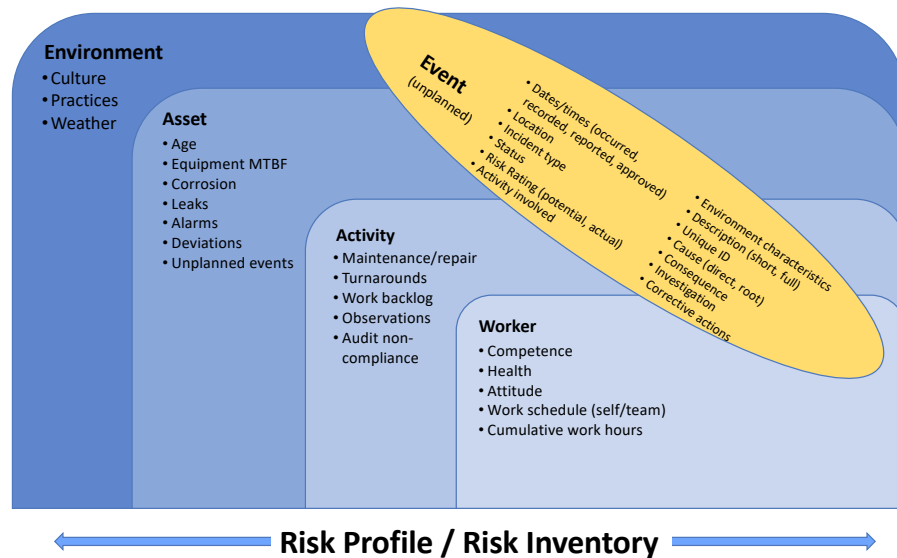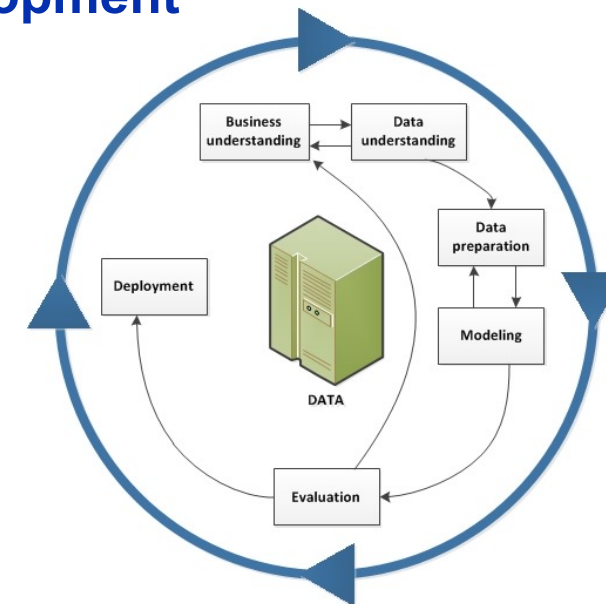
1. Phase of Work

Plan ➡ Prepare ➡ Execute ➡ Lookback

2. Depth of Insight

Automated Actions

Context Driven Recommendations

Event Driven Notifications

Pre-defined Reports & Analysis

Ad-hoc / Manual Inquiry

Weeks/Days ——————————▶ Minutes/Seconds

3. Breadth of Data

Environment
• Culture
• Practices
• Weather

Asset
• Age
• Equipment MTBF
• Corrosion
• Leaks
• Alarms
• Deviations
• Unplanned events

Activity
• Maintenance/repair
• Turnarounds
• Work backlog
• Observations
• Audit non-compliance

Worker
• Competence
• Health
• Attitude
• Work schedule (self/team)
• Cumulative work hours

Event (unplanned)
• Dates/times (occurred, recorded, reported, approved)
• Location
• Incident type
• Status
• Risk Rating (potential, actual)
• Activity involved
• Environment characteristics
• Description (short, full)
• Unique ID
• Cause (direct, root)
• Consequence
• Investigation
• Corrective actions

◀—— **Risk Profile / Risk Inventory** ——▶

4. Model development

Business understanding ⇄ Data understanding
Data preparation
Modeling
Evaluation
Deployment
DATA

IBM

The first three models work together to establish context for a single model or a program managing hundreds of models.

Phase of Work
(PoW)

Breadth of Data
(BoD)

Depth of Insight
(DoI)

IBM
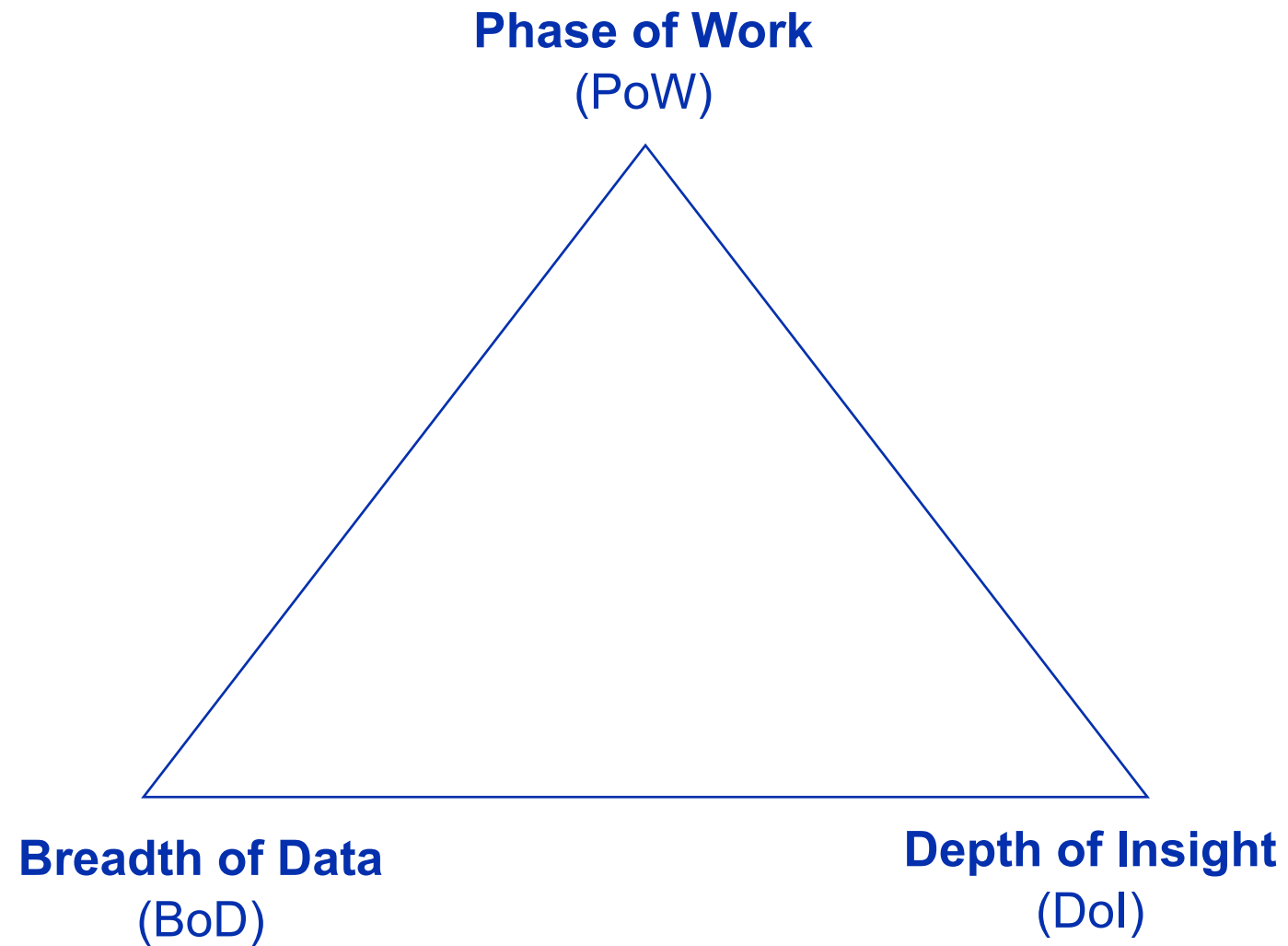
Phase of Work describes the specific business context of the work to be performed from planning through execution plus lookbacks.
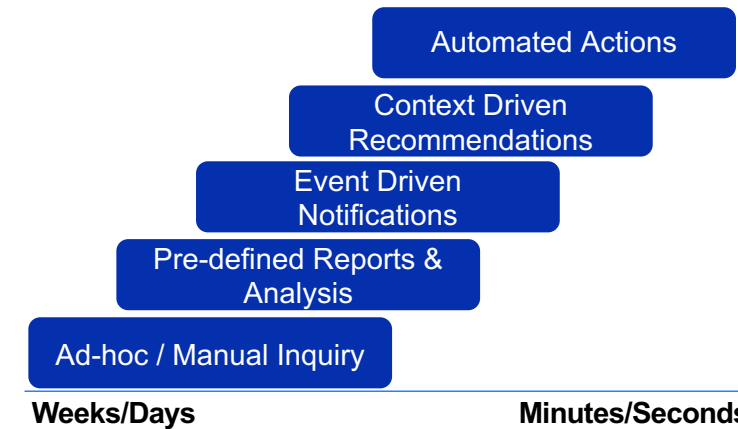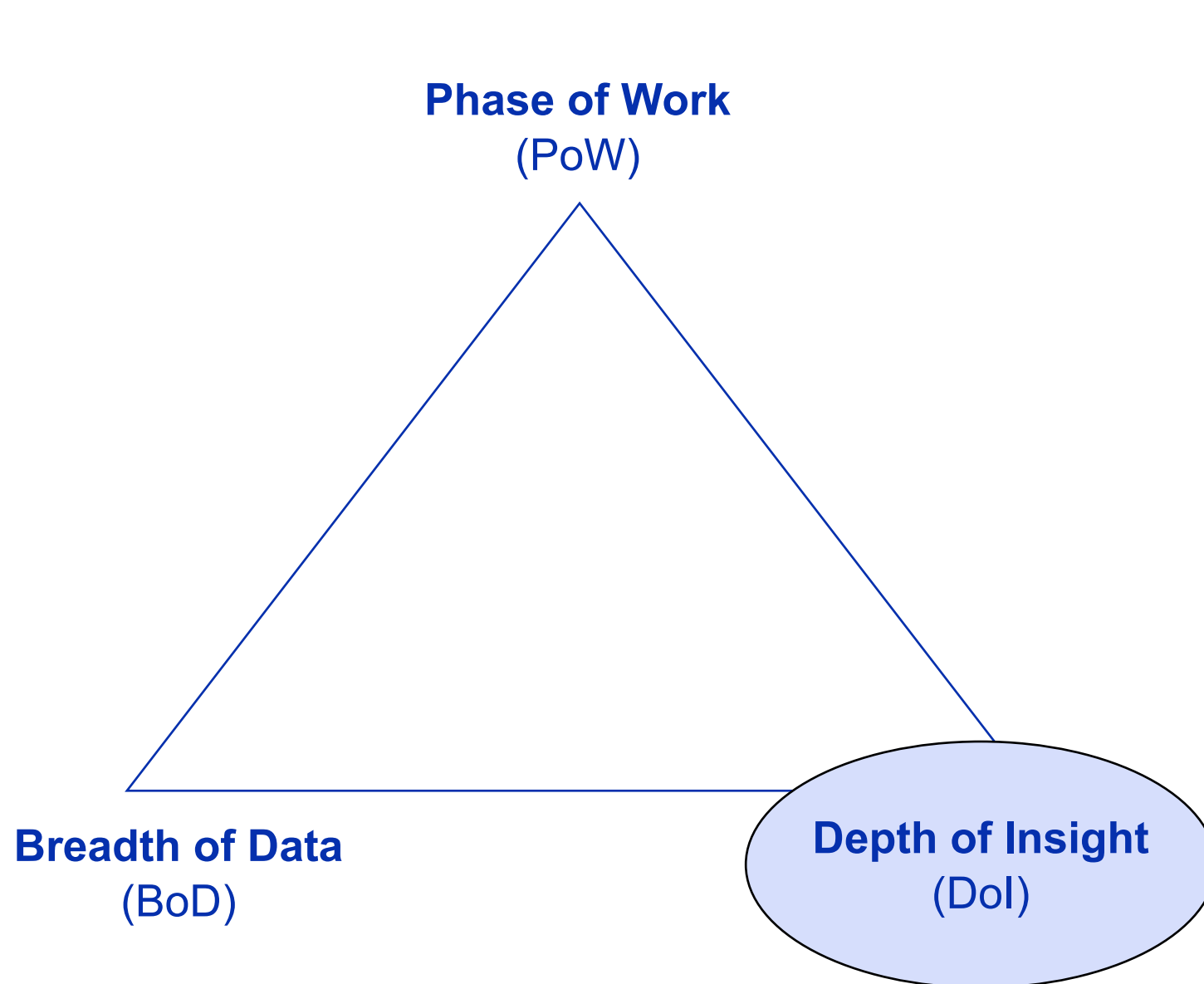
Phase of Work
(PoW)

Breadth of Data
(BoD)

Depth of Insight
(DoI)

Plan ➡ **Prepare** ➡ **Execute** ➡ **Lookback**
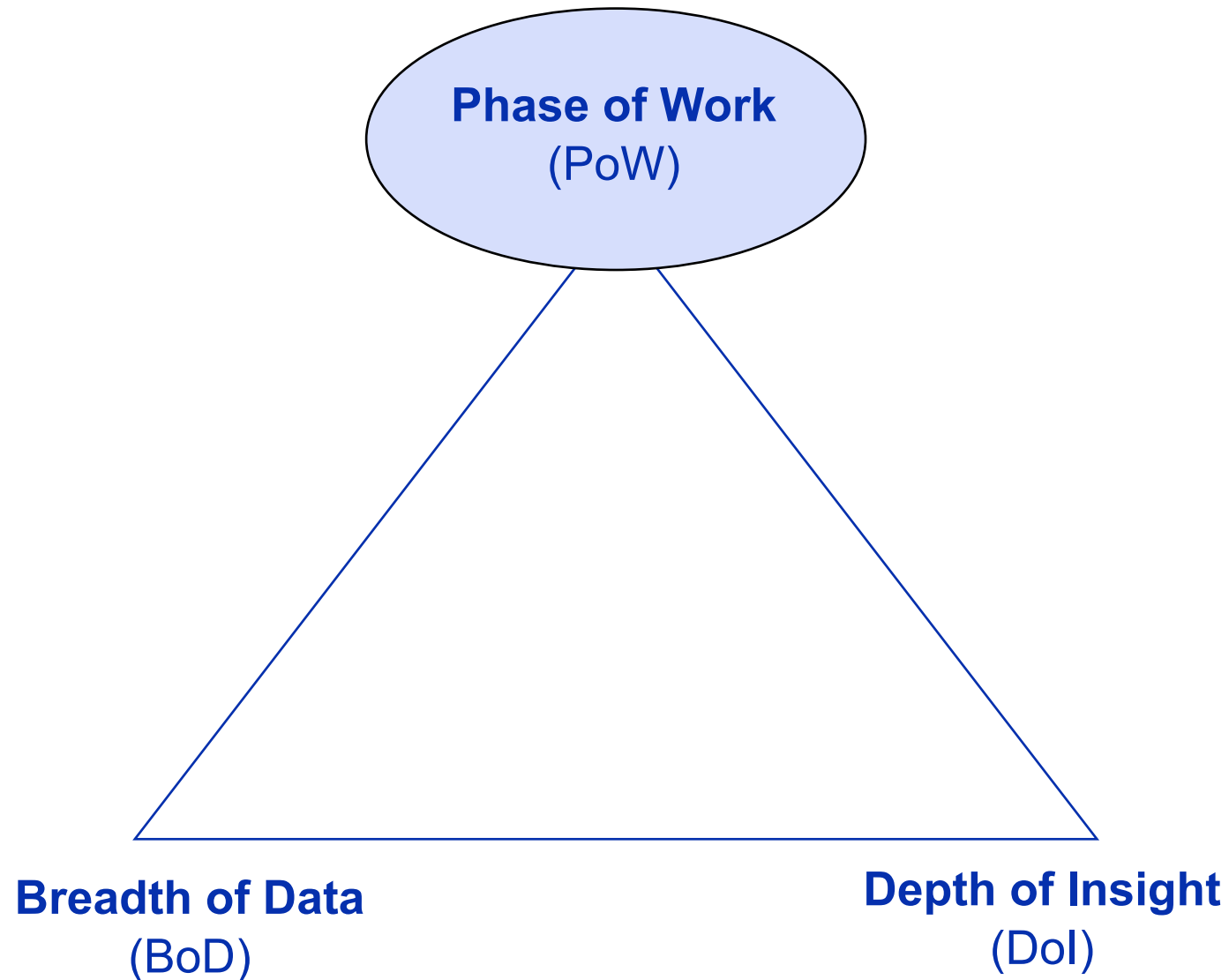
Phase of Work
(PoW)

Phase of Work (PoW) describes the specific business context of the work to be performed from planning through execution plus lookbacks. PoW is critical to estimating the business impact and benefits from the effort to develop and maintain the model(S).

Common Values:
- Planning
- Preparation
- Execution
- Look Back

IBM

Depth of Insight indicates overall complexity. Dimensions include user experience, data frequency, integration, model complexity, automation.

Phase of Work
(PoW)

Breadth of Data
(BoD)

Depth of Insight
(DoI)

Automated Actions

Context Driven
Recommendations

Event Driven
Notifications

Pre-defined Reports &
Analysis

Ad-hoc / Manual Inquiry

Weeks/Days   **Minutes/Seconds**

Depth of Insight
(DoI)

Depth of Insight (DoI) indicates overall complexity. Dimensions include user experience, data frequency, integration, model complexity and automation. Recommendations or automated actions can require a lot of contextual data vs a predictive model indicating the likelihood of an event.

Common Values:
- Ad-hoc / Manual Inquiry,
- Pre-defined Reports & Analysis,
- Event Driven Notifications,
- Context Driven Recommendations,
- Automated Actions

IBM

Breadth of Data identifies all the types of data relevant for making a business decision. Going beyond event data provides more context.

Phase of Work
(PoW)

Breadth of Data
(BoD)

Depth of Insight
(DoI)
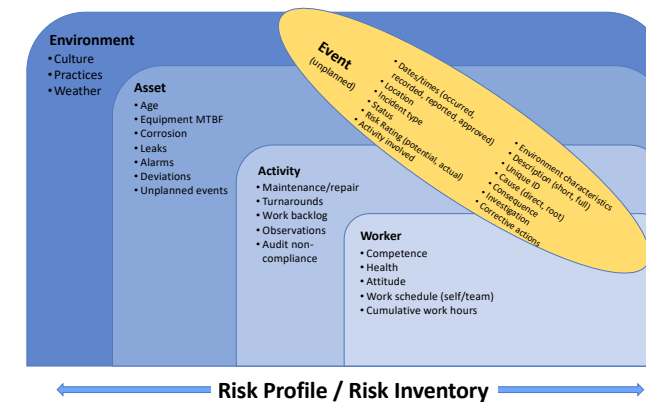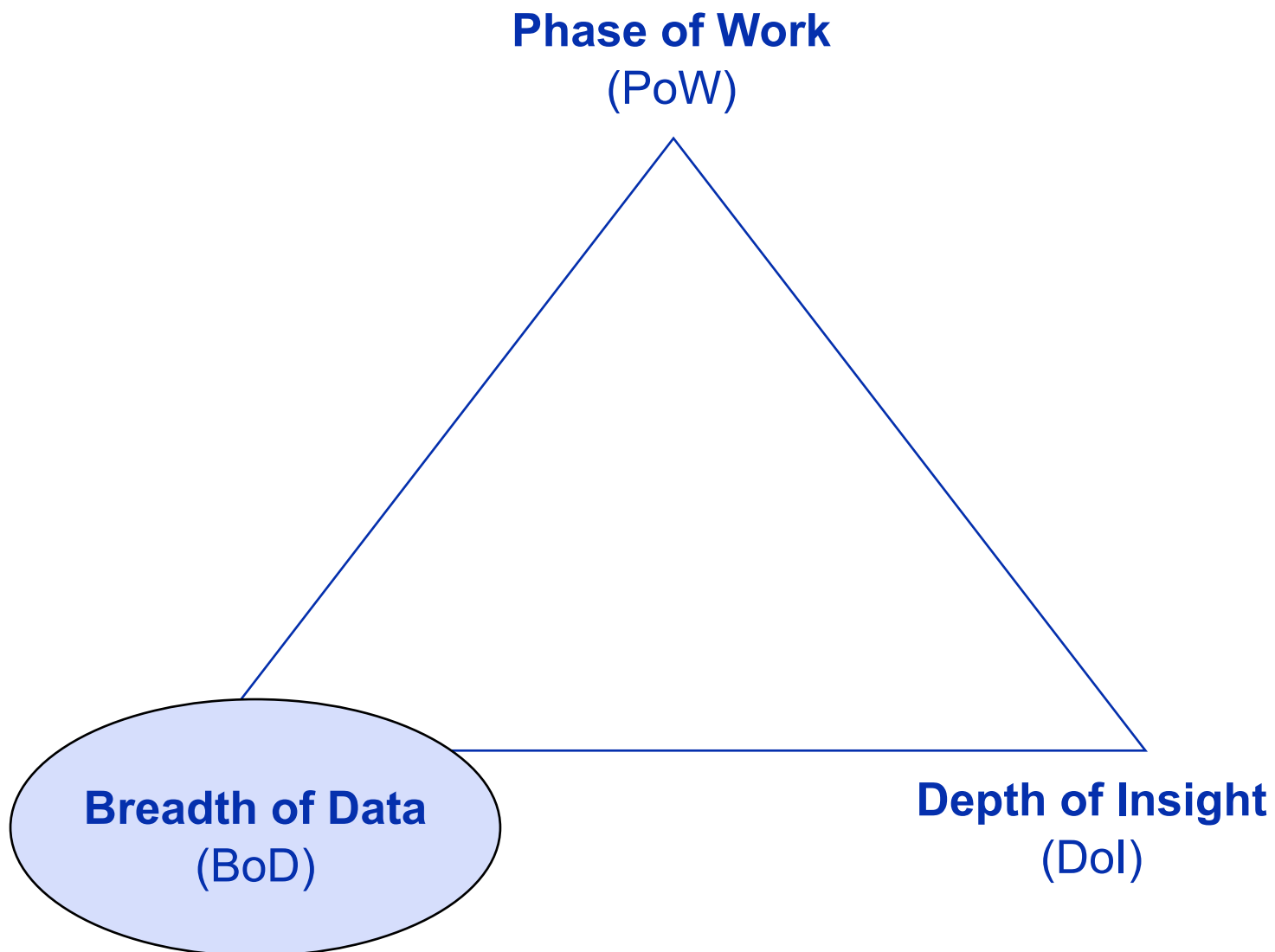
Breadth of Data
(BoD)

Breadth of Data (BoD) identifies all the types of data relevant for making a business decision. Going beyond event data provides more context. Event data tends to be limited in what is captured and very little about what proceeded the event. Worker, activity, asset and environmental data gives a holistic view of all potential factors.

Common Values:
- Environment
- Asset
- Activity
- Worker
- Event

IBM

After defining the basic business context, insight required and data to enable, you're ready to start the Model Development lifecycle.
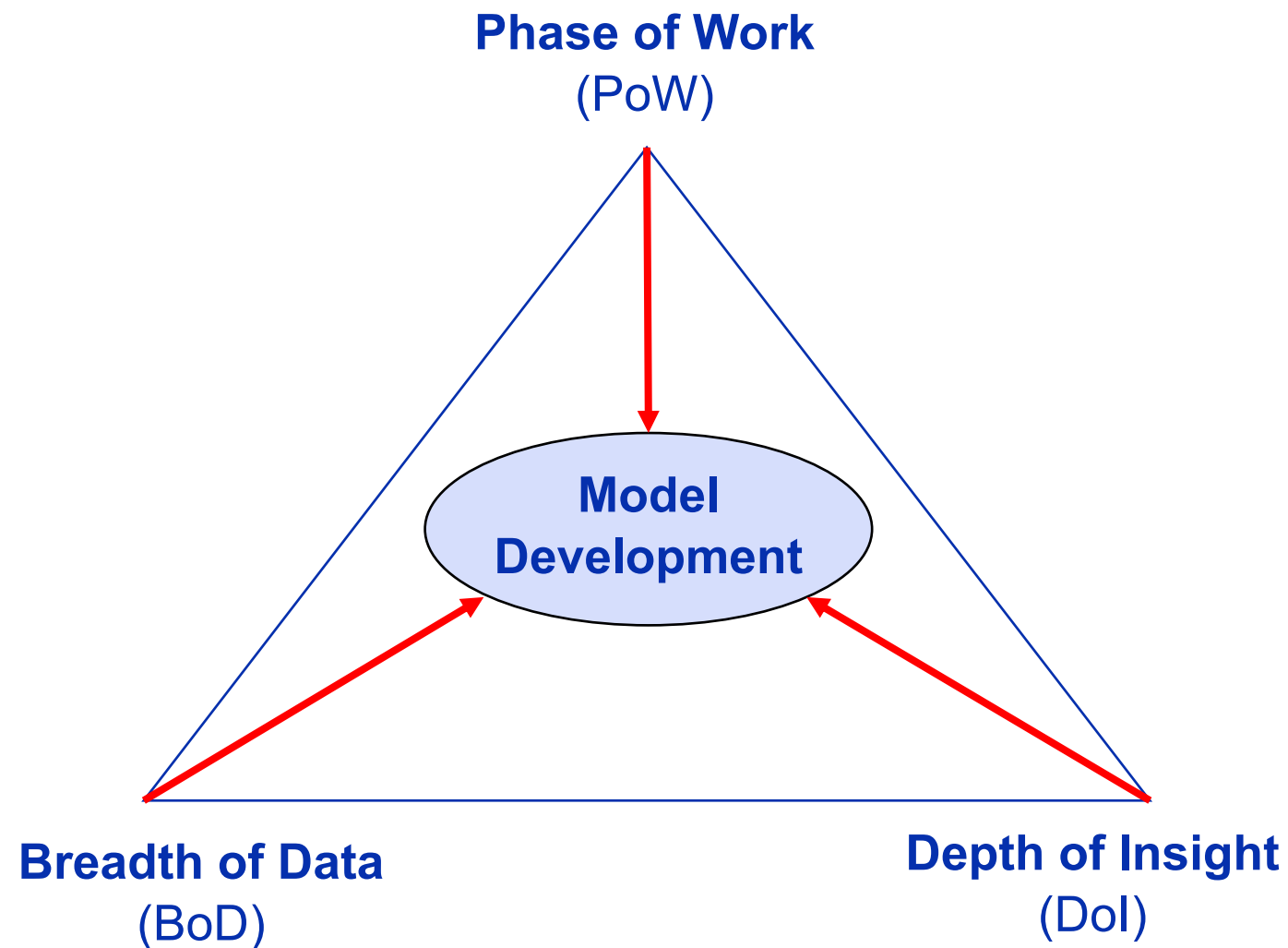
Phase of Work
(PoW)

Model Development

Breadth of Data
(BoD)

Depth of Insight
(DoI)

Model Development is the continuous process of planning, developing and deploying models (e.g. CRISP-DM.)

Phase of Work
(PoW)
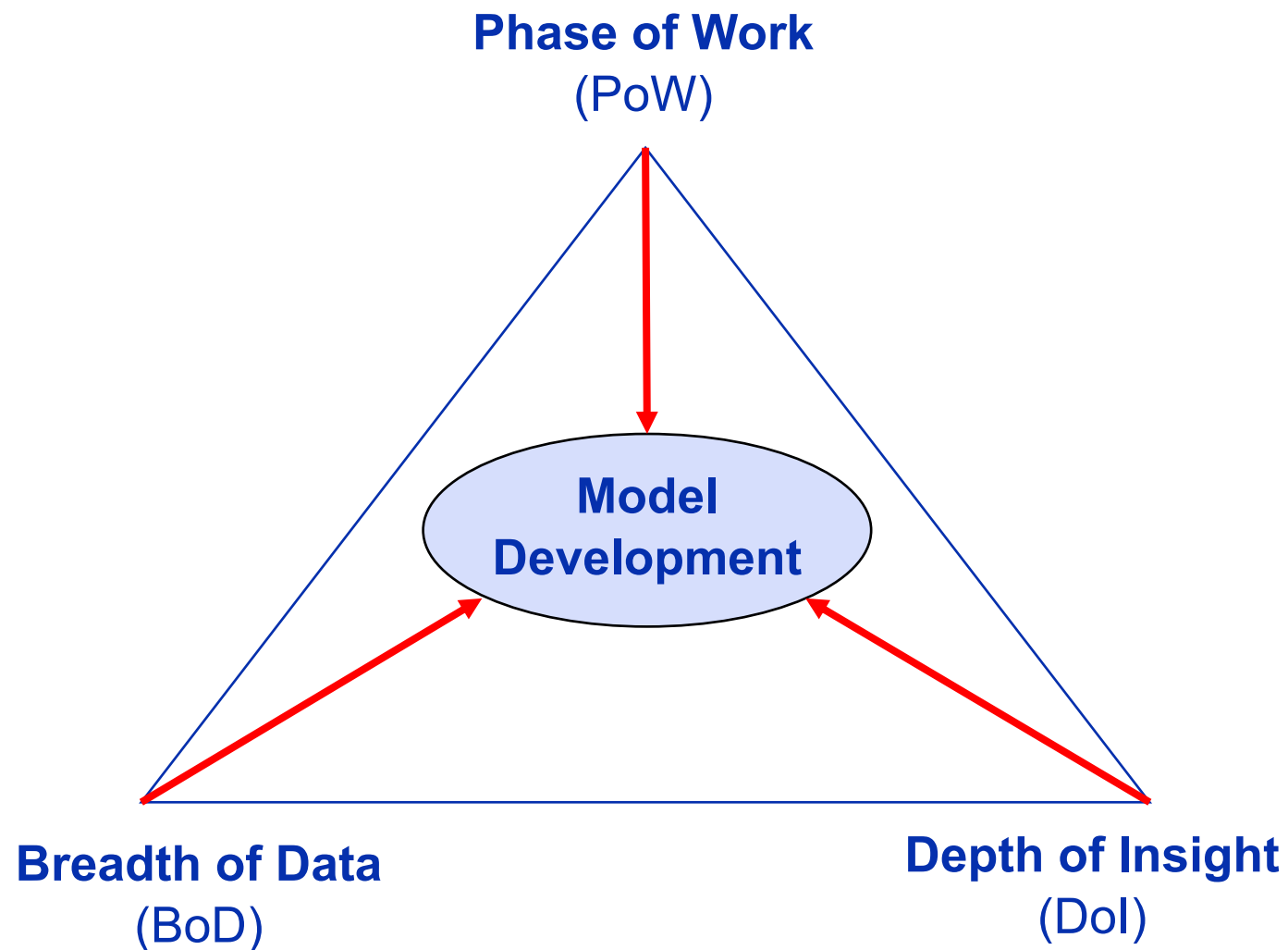
Model Development
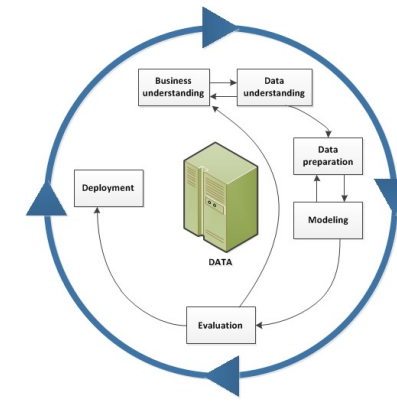
Model Development

Model development is the continuous process of planning, developing and deploying models (e.g. CRISP-DM.) The *Data Preparation* stage can frequently consume up to 80% of the total project effort. The Modeling phase, arguably the most exciting, is usually the shortest.

Breadth of Data
(BoD)

Depth of Insight
(DoI)

Common Values:
- Business Understanding
- Data Understanding
- Data Preparation
- Modeling
- Evaluation
- Deployment

IBM

Real world success requires a deep understanding of the business process and desired outcomes.

Solution		Phase of Work	Depth of Insight	Breadth of Data
Multi-plant optimization for continuous operations	Runs 100+ models simultaneously to address potential process upsets (flag, response, recovery) and opportunities to increase profitability (identify, recommend)	• Plan • Execute	• Context Driven Recommendation • Event Driven Notification	• Asset performance • Operational Constraints • Market pricing
Exploration drill target selection	Predict the presence of economic minerals (e.g. gold, nickel) with confidence score within a 3D space. Site-wide 3D geologic query engine.	• Plan • Prepare • Execute	• Context Driven Recommendation • Ad-hoc/Manual Query	• Drill holes w/intervals • Block models • Geological models • …
Predictive Drilling	Identified drilling events (e.g. pack-off) and integrated model (i.e. ensemble model) combining a traditional predictive model and cognitive analytics (unstructured information.) Unstructured information improves prediction sensitivity and provides earlier warnings than structured data alone.	• Execute	• Event Driven Notification • Context Driven Recommendation	• Sensor data (hook load, depth, RPM,…) • Rig state • Drill reports

IBM

Often multiple models are running in parallel or even in concert for a single recommendation or insight.

Solution		Phase of Work	Depth of Insight	Breadth of Data
Haul Truck *Predictive Maintenance*	Developed a suite of 75+ models for predicting multiple types of truck failures and optimizing maintenance schedule based on weather impacts. Optimizations included haul road optimization, payload optimization and turn analysis and operator, crew & site analytics	• Plan • Prepare • Execute	• Context Driven Recommendation • Event Driven Notification • Pre-defined Reports & Analysis	• Equipment sensors • Fluid analysis • Mobile inspections
Haul Truck *Site Optimization*		• Plan • Prepare • Execute	• Context Driven Recommendation	• Fleet Management • ERP • Weather • Team performance
Pipeline Corrosion Management	Applied deep learning (convolutional neural networks) and visual recognition to pipeline inspection images (i.e. magnetic flux) to automate the identification of pipe wall degradation or leaks.	• Plan	• Pre-defined Reports & Analysis	• Pipeline Images

IBM

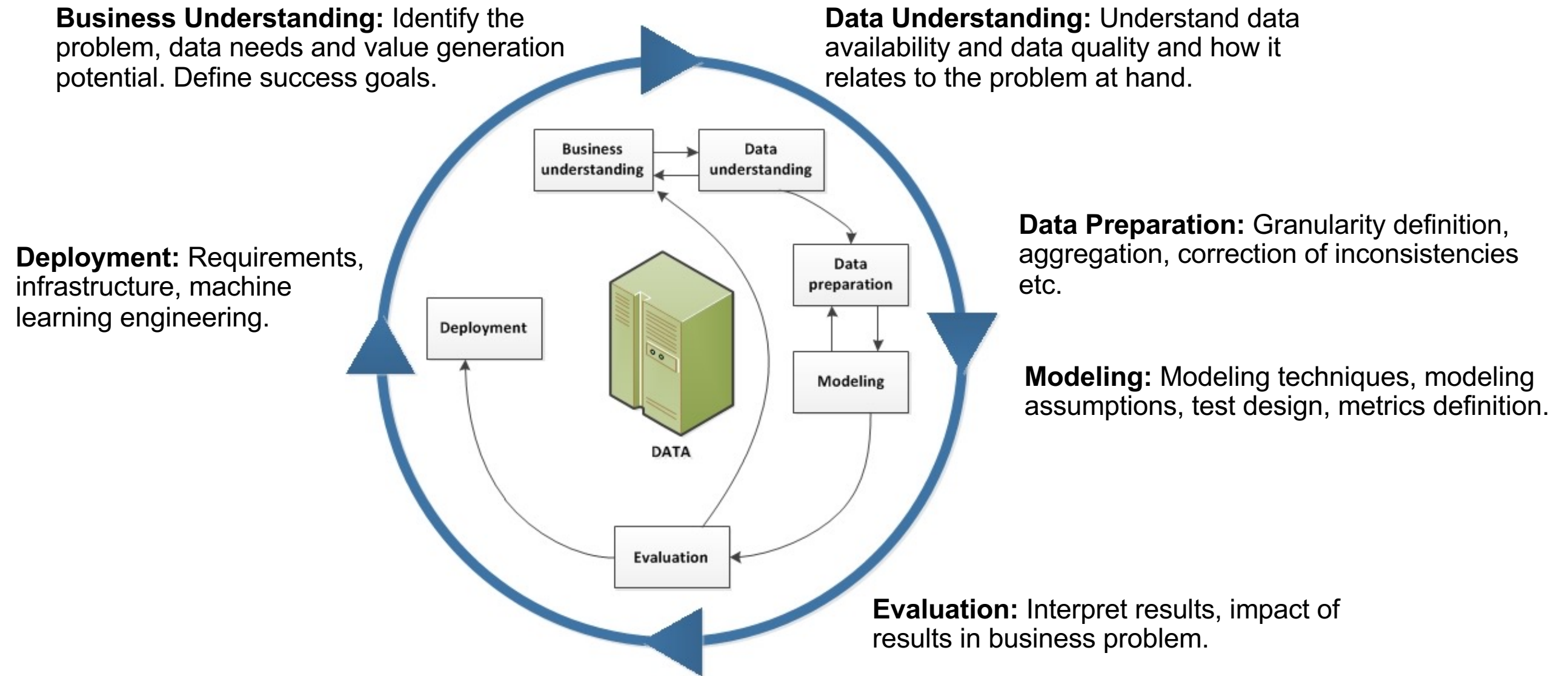# Health and Safety is better with AI

Assisted
Intelligence

Augmented
Intelligence

Autonomous
Intelligence

CRISP-DM as a framework for model development.

Business Understanding: Identify the problem, data needs and value generation potential. Define success goals.

Data Understanding: Understand data availability and data quality and how it relates to the problem at hand.

Deployment: Requirements, infrastructure, machine learning engineering.

Data Preparation: Granularity definition, aggregation, correction of inconsistencies etc.

Modeling: Modeling techniques, modeling assumptions, test design, metrics definition.

Evaluation: Interpret results, impact of results in business problem.

CRISP-DM, which stands for *Cross-Industry Standard Process for Data Mining*, is an industry-proven way to guide your data mining efforts.

Business Understanding
Health and Safety requirements

**Phase of Work?
Depth of Insight?
Breadth of Data?**

Clear definition of the problem to be solved
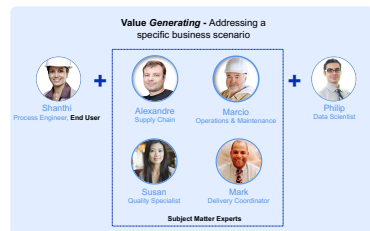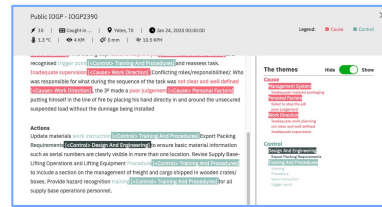
SME + Data Scientist + User

Many iterations

IBM
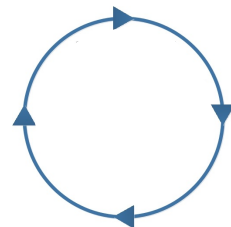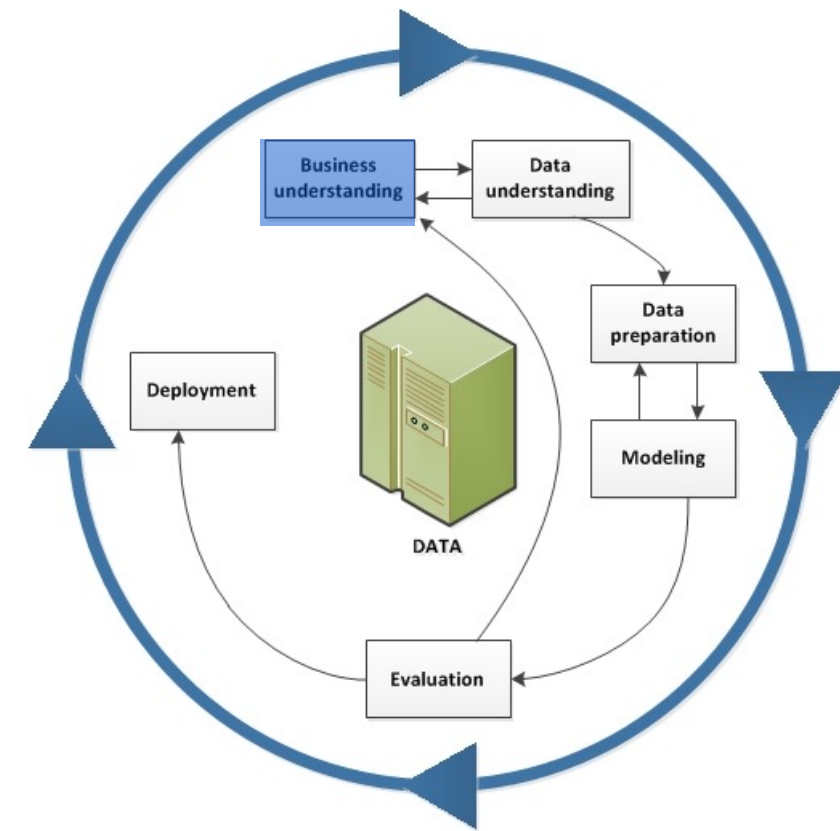
Data Understanding & Data Preparation
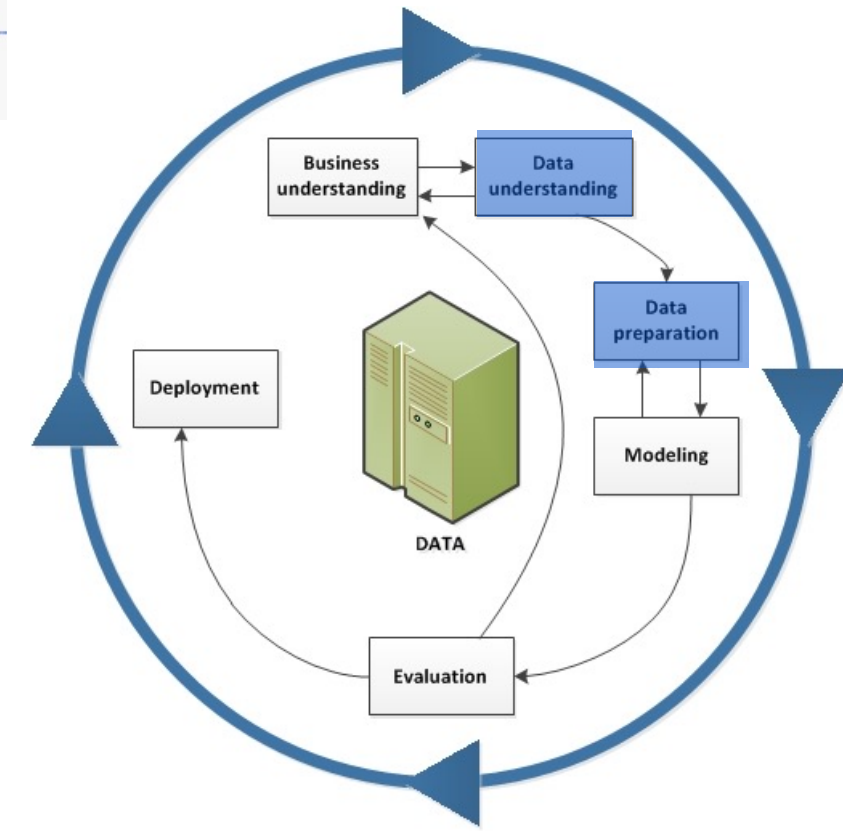## Selecting data and understanding data useability

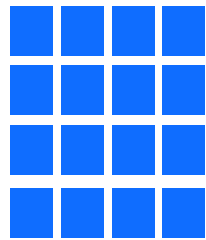**Structured**

Unstructured
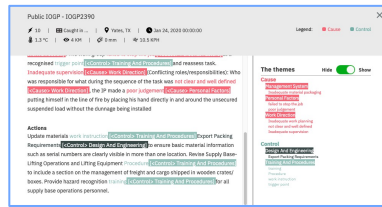
Image

Video

Audio

IBM
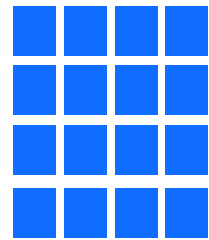
Modeling
Selecting and training a model(s)

Structured   **Unstructured**   **Image**   **Video**   **Audio**

|		
Supervised	Regression	Neural Network Regression Support Vector Machine Decision Trees Linear regression
	Classification	Neural Network Classification Support Vector Machine Decision Trees
Unsupervised	Clustering	K-means DBSSAN Agglomerative Clustering
	Natural Language Processing	Named entity recognition Summarization Natural Language Understanding
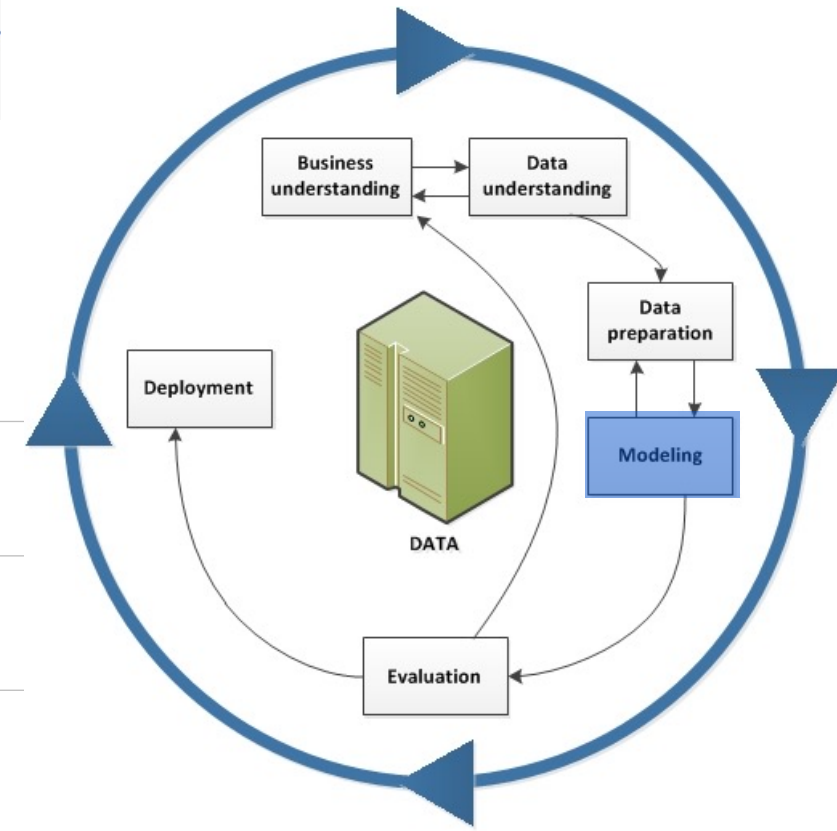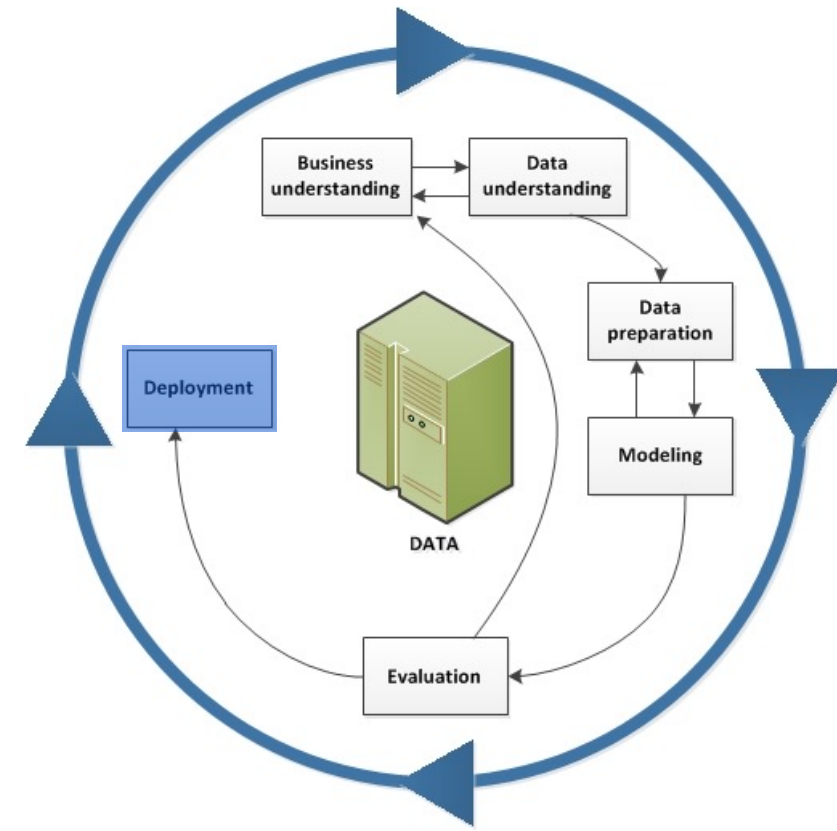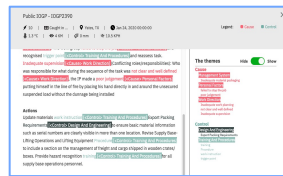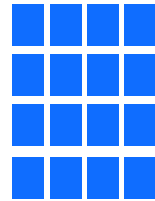
IBM

Deployment
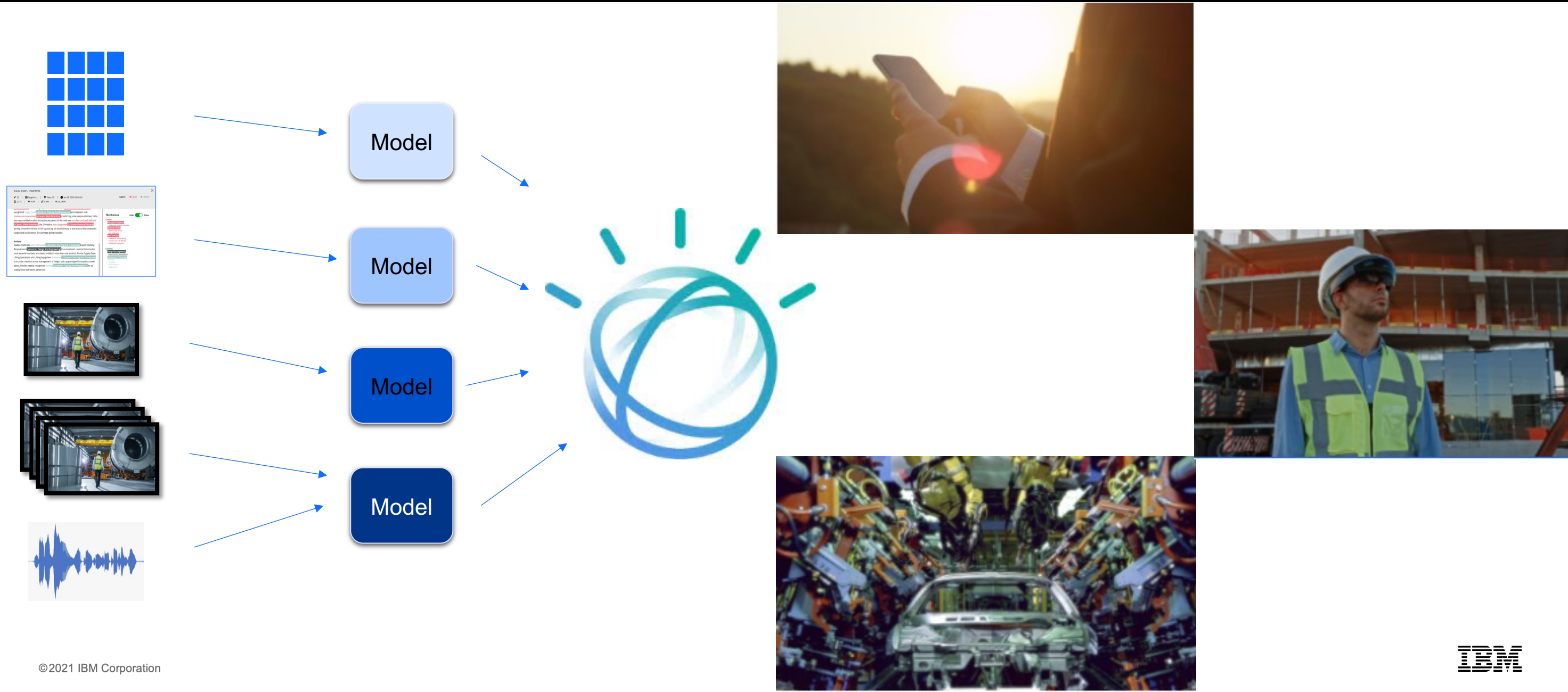## Deploying and managing models over time

Deployment
Deploying and managing models over time
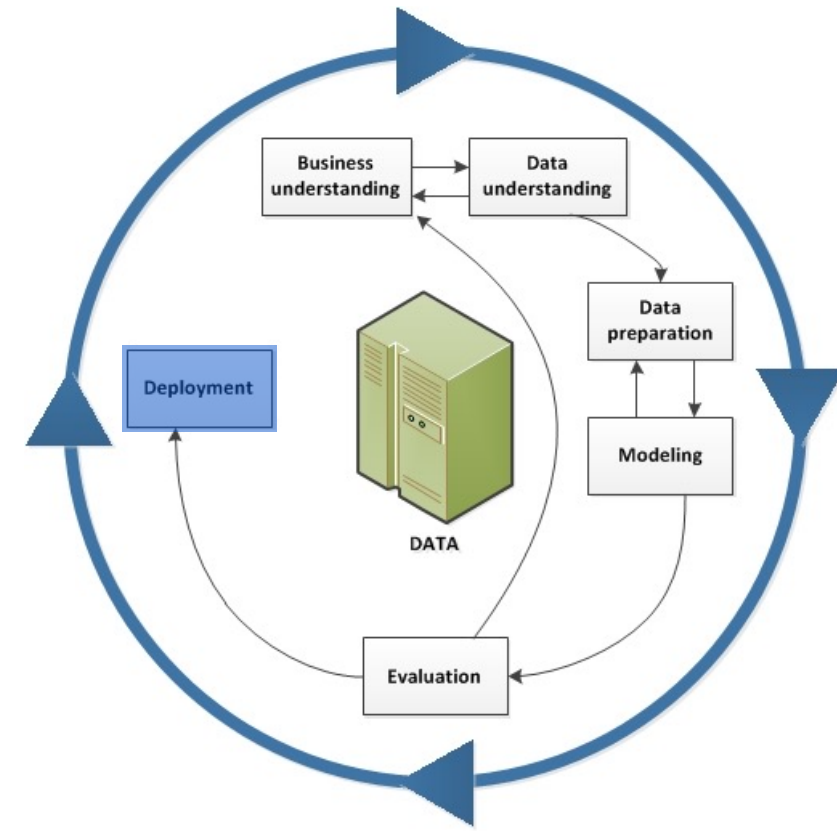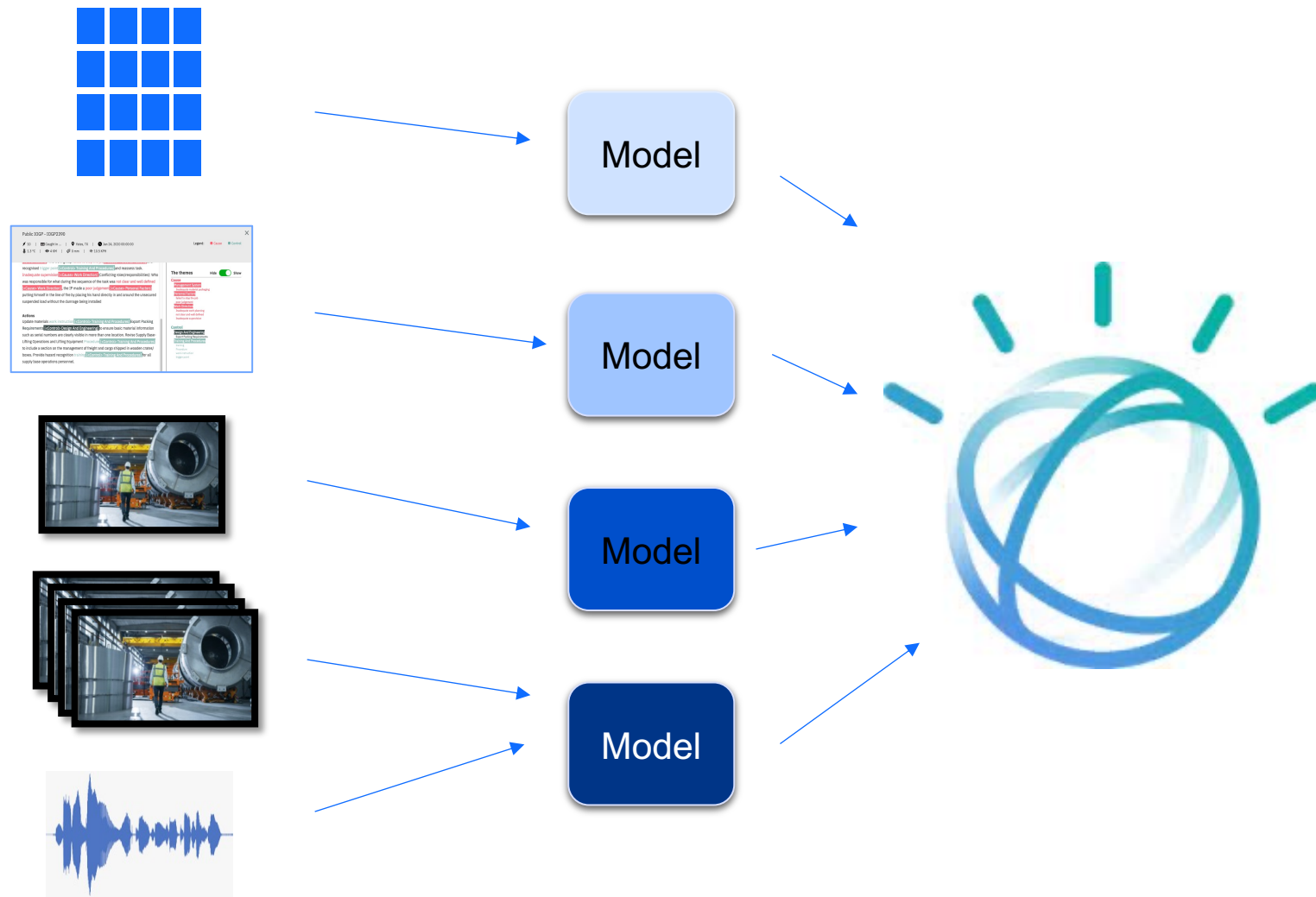
Deployment
Deploying and managing models over time

We're at an inflection point as demand for health & safety data continues to grow and data science tools provide low code or no code interfaces making it easy generate new insights using a wide range of techniques!

What you learned

- Key participants in developing data science based models (slide 9-10)

- Frameworks for assessing individual or 100's of use cases (slides 11-19)

- Key data domains to consider for health & safety analysis (slide 15)

- Approach for designing, building/training machine learning models (slide 21)

- Common data formats (slide 23)

- Common data analysis techniques (slide 24)

- Where to learn more about everything above! (slide 29)

- Lessons learned designing, building/training machine learning models in health & safety (Throughout!)

IBM

Where to learn more and earn some street cred!

General AI/ML

- McKinsey & Co. *Executive's guide to developing AI at scale* - https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/executives-guide-to-developing-ai-at-scale#intro

- McKinsey Analytics – multiple articles on AI, Data Strategy and Organization - https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights

- *The Ultimate guide to AI, Data Science & Machine Learning, Articles, Cheatsheets and Tutorials ALL in one place* - https://www.linkedin.com/pulse/all-cheatsheets-one-place-vipul-patel/

Coursera Certificates & Badges

IBM Data Science Professional Certificate - https://www.coursera.org/professional-certificates/ibm-data-science

Data Science Orientation - https://badges.mybluemix.net/badge/6647ce07-7fa3-40b9-91dc-d753803b3c22

Artificial Intelligence Essentials - https://badges.mybluemix.net/badge/bfceb0d1-44bd-49de-9ac5-46090839ff20

Artificial Intelligence Foundations Specialization - https://badges.mybluemix.net/badge/a00ba7ba-226c-4dfb-8c0b-63f61bc21e87

IBM AI Foundations for Business Specialization - https://badges.mybluemix.net/badge/f2166942-64f6-4d68-9e63-9e7280681e7e

IBM

Let's connect and continue the conversation!!!

Scott Kimbleton | IBM

Global AI Solutions, Deployment Leader

Energy & Natural Resources

Phone: (919) 247-3793

Email: Kimbleton@us.ibm.com

Linkedin: https://www.linkedin.com/in/scottkimbleton/

Graziella Caputo | IBM

Sr. Data Scientist & Manager

Data Science Elite

Email: Graziella.Caputo@ibm.com

Linkedin: https://www.linkedin.com/in/graziellacaputo/