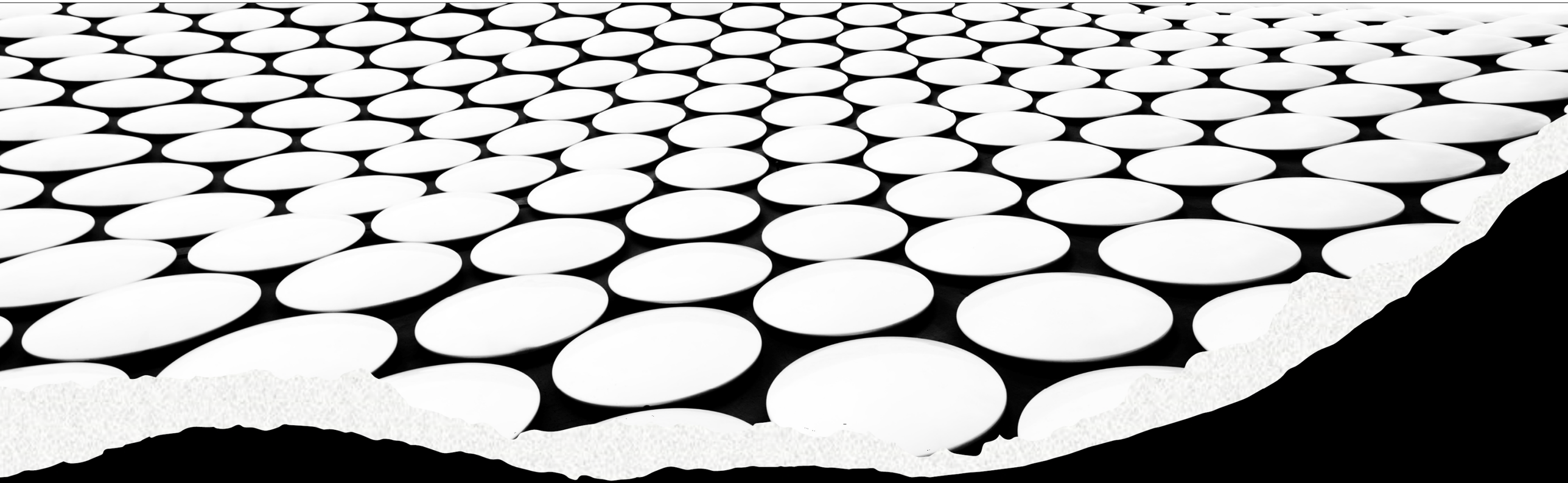


Governing Your Big Data

Meredith K. Cebelak, PhD, PE
Gresham Smith

#GreshamSmith

Agenda

- Big Data
- What is the Issue?
- Governing It
- How Do We do This?

#GreshamSmith

Big Data

What is your Big Data Experience?

Big Data??

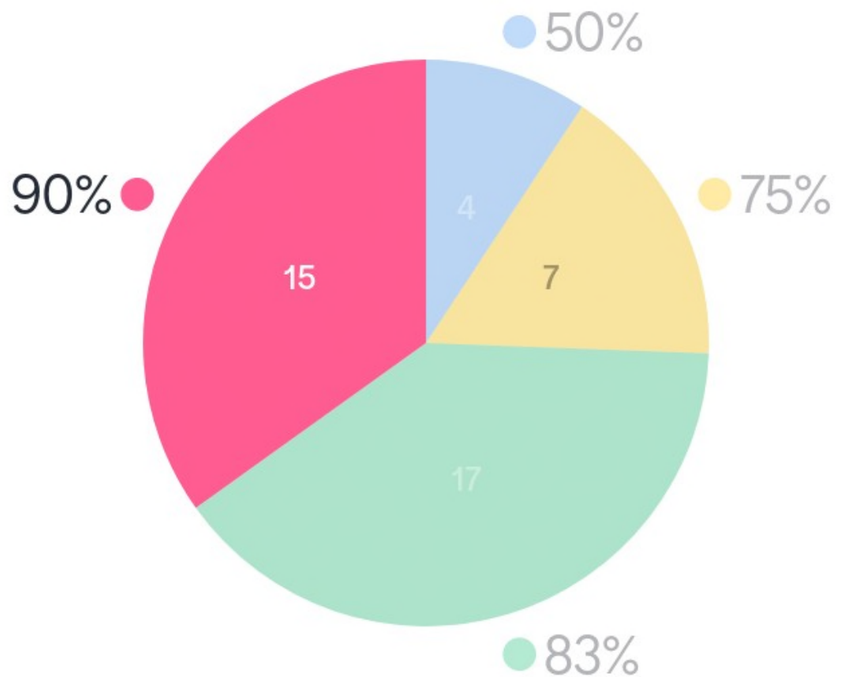
Novice Here

I have Played Around A Little

I've Gone Pro


DATA is EVERYTHING...
EVERYTHING is DATA

In the last 2 years how much data has been created?

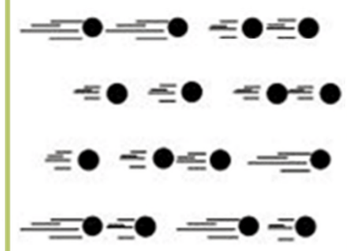
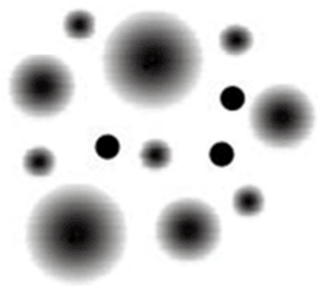


#GreshamSmith

Data – A Modern Currency

- In Last 2 Years, 90% of Data has been Created
- IoT is Increasing the Data Created
- By 2025, 41.6 Billion Connected IoT Devices Generating 79.4 ZB of Data

Data – A modern Currency

Volume	Velocity	Variety	Veracity	Value
				
Data at Rest	**Data in Motion**	**Data in Many Forms**	**Data in Doubt**	**Data into Money**
Terabytes to Exabytes of existing data to process	**Streaming data, requiring milliseconds to seconds to respond**	**Structured, unstructured, text, multimedia,…**	**Uncertainty due to data inconsistency & incompleteness, ambiguities, latency, deception, model approximations**	**Business models can be associated to the data**

Adapted by a post of Michael Walker on 28 November 2012

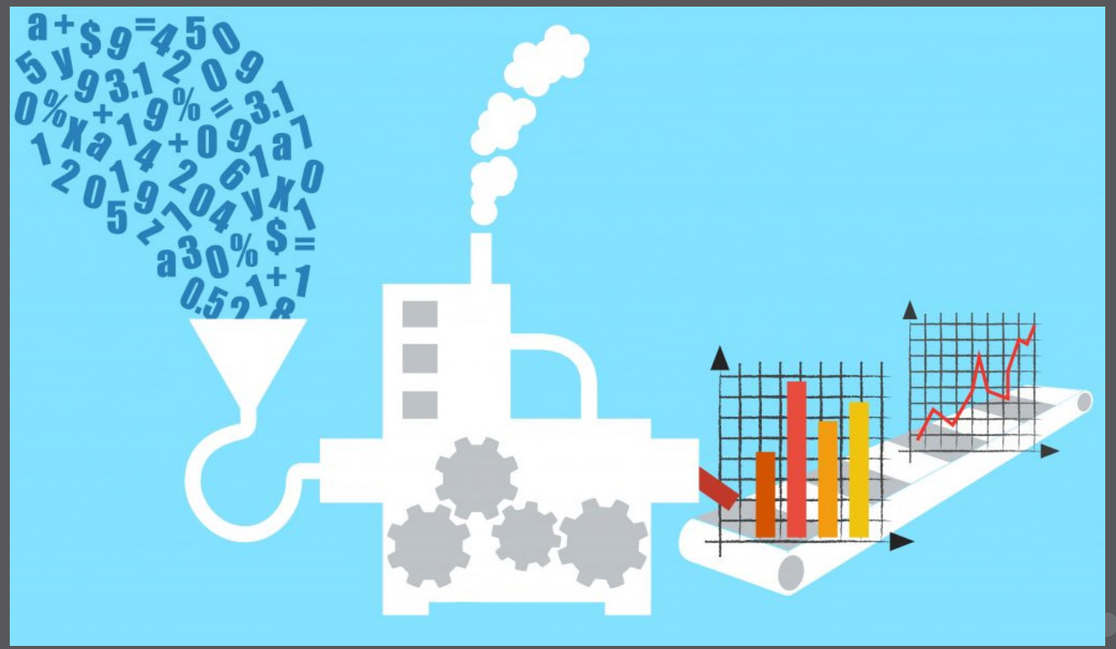
Data - Transportation

- **Roadway Sensing – GPS, Cellular, Radar, Video, Bluetooth**
- **Roadway Assets – Pavement Conditions, Guardrail, Signage, Traffic Signal Poles, Lighting Systems**
- **Workforce – Emails, Institutional Knowledge, Operations, Maintenance Records**

What is the Issue?

#GreshamSmith

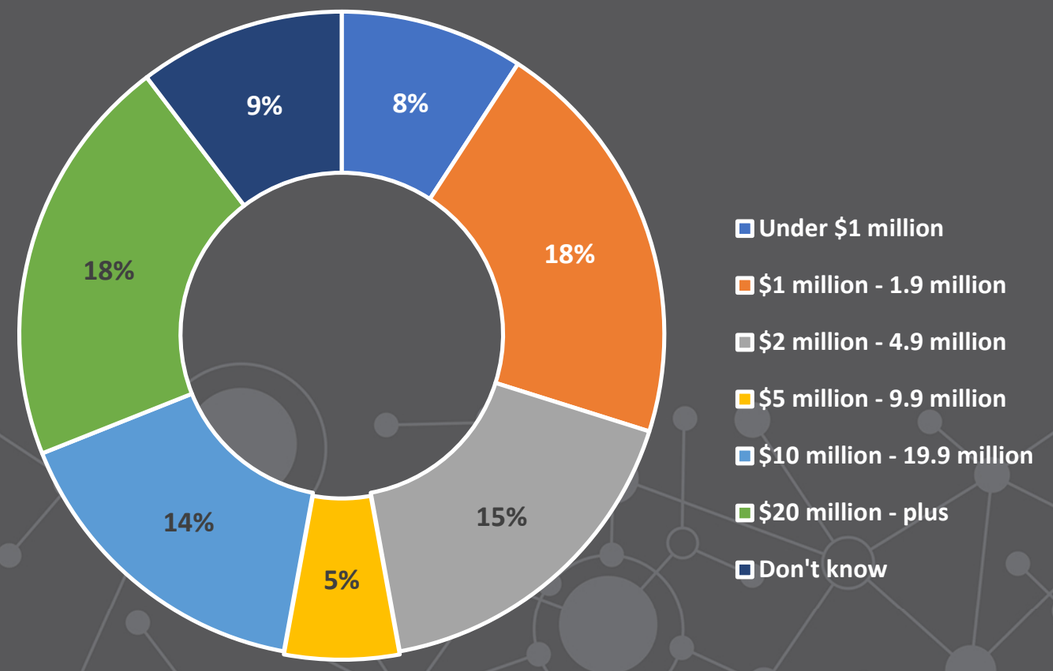
Are All Data Good Data?



#GreshamSmith

How big is the Problem?

- *According to Forbes, Data-Related Problems Cost the Majority of Companies More Than* ***$5 million Annually…***



Common Challenges

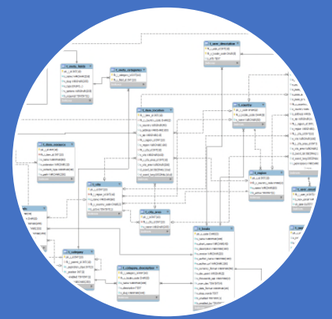
#GreshamSmith

Person to Person Sharing

Reliance on Getting Data from Individuals, Not from Applications and Reporting Tools

Extensive Manual Processing

Lots of Manual, Home Grown Processes for Copying & Transferring Data



Limited or No Standardization

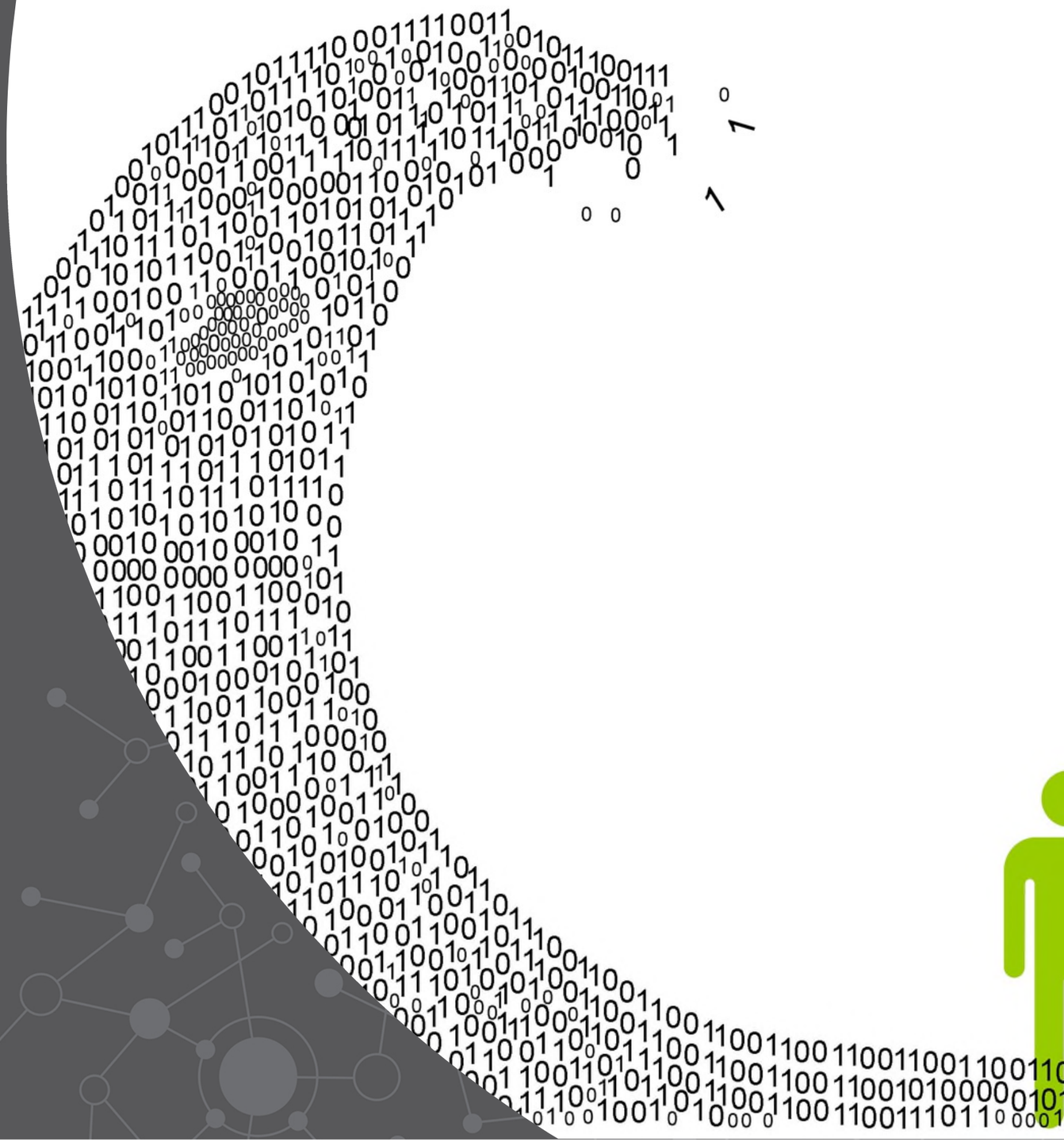
Lots of Effort to Relate Information from Multiple Data Sources

Lack of Sustainability

Data Driven Agency = New Challenges

IT IS A STRUGGLE OUT THERE!

- Too Much Data
- Knowing What Can be Done with Data
- Only Use about 10% of Data Generated
- AI and Machine Learning

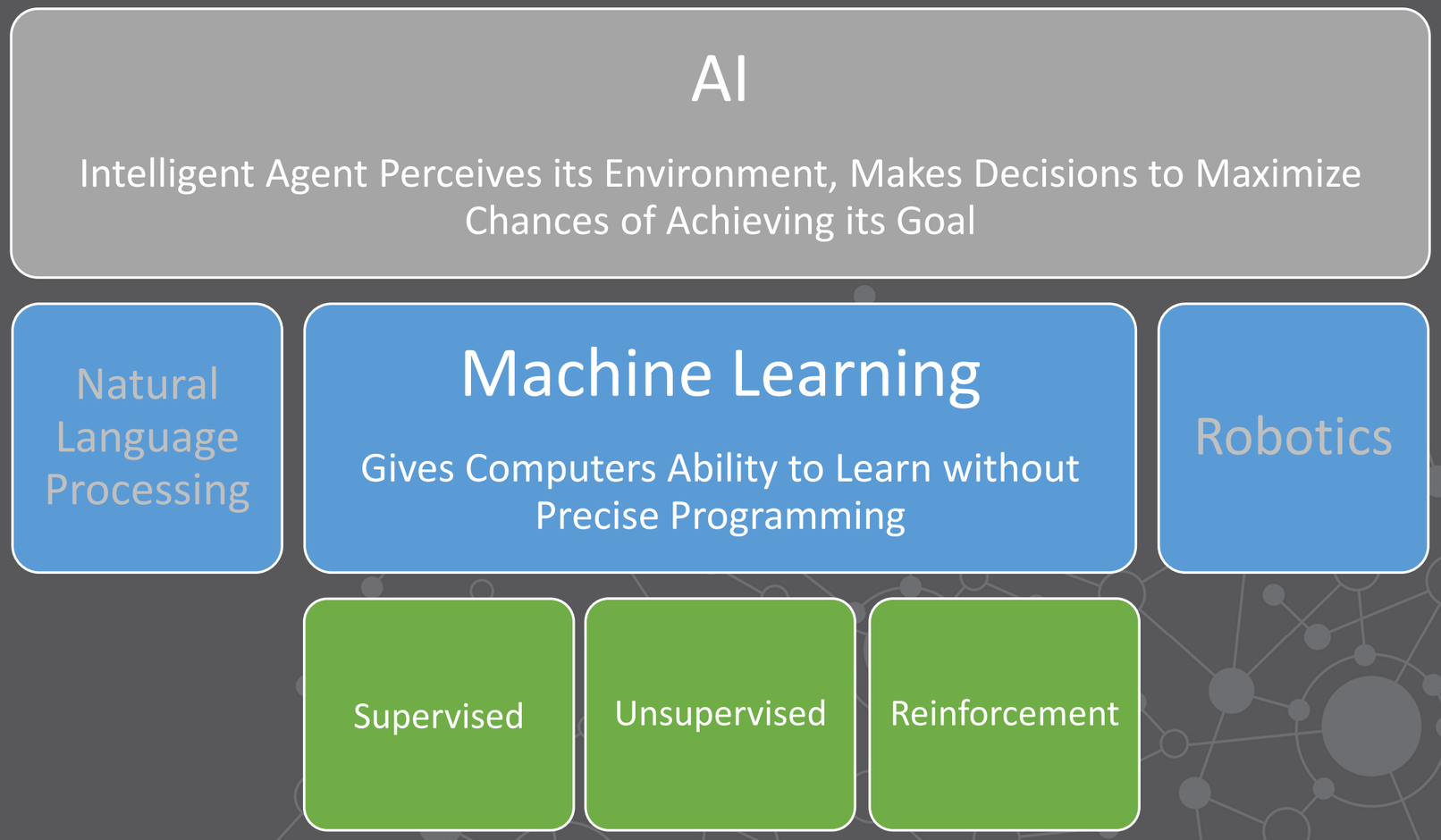


What AI or Machine Learning are you familiar with?

#GreshamSmith

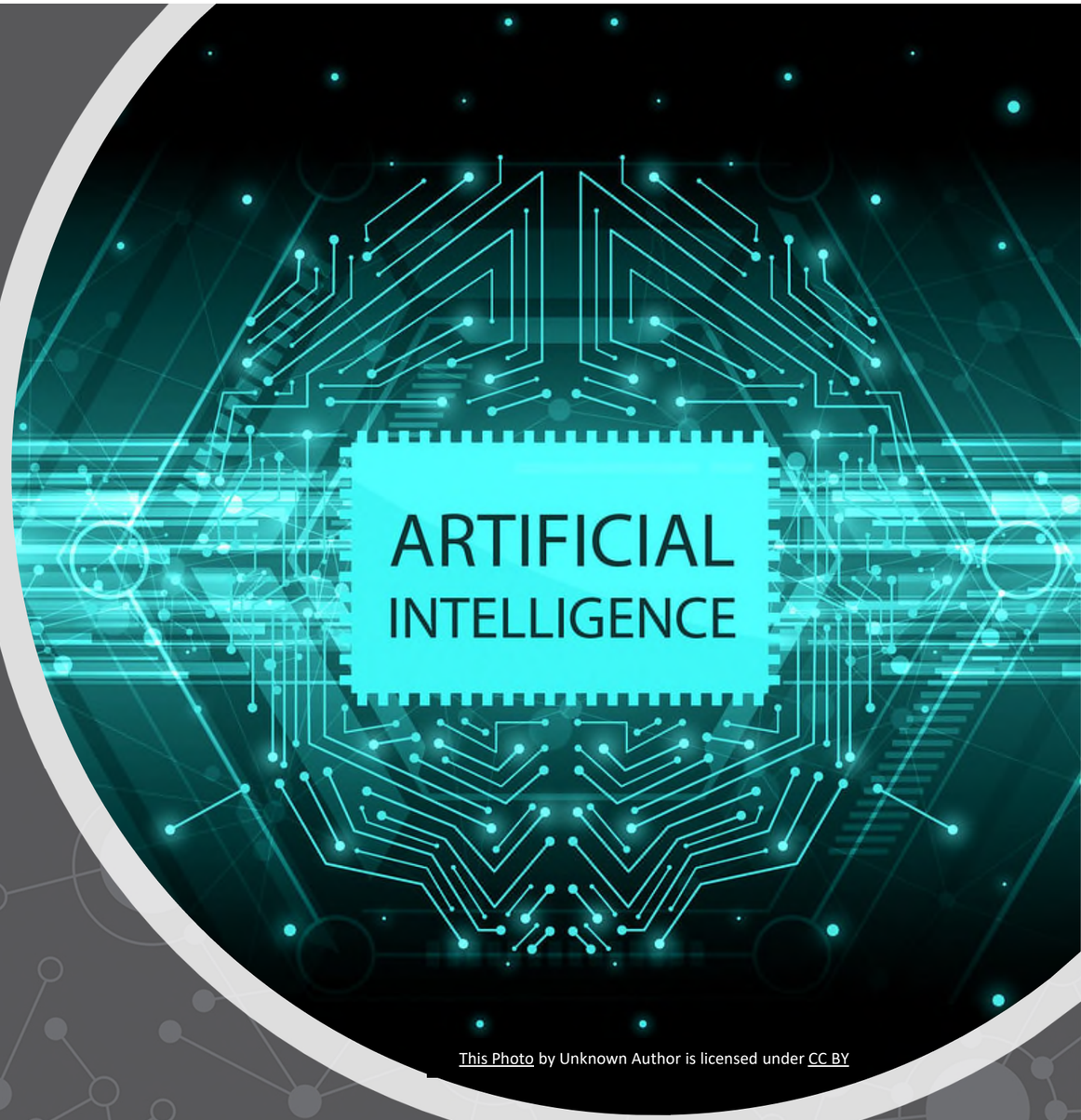
AI and Machine Learning

- Limited Background In or Exposure to AI
- Blackbox Experience – Some Algorithm is Used, Gives “An Answer”



AI and Machine Learning - Existing Transportation Applications

- Chatbots and QA Systems - Enables New Insights into Data
- Neural Networks - Analyze Imagery from Many Sources for Incident Detection, Incident Management, and Traffic Data Collection
- Fuzzy Logic - Being Used by DOTs for Ramp Metering, May have Applications for the “If…Then” Rule for Decision-Support Systems
- Unsupervised AI Systems - Learn New Ways to Control Traffic and Coordinate Integrated Corridor Management Actions Across Control and Advisory Technologies
- Driverless Vehicles and Unmanned Aerial Systems (UAS) - Improve TSMO Staff Safety and Productivity

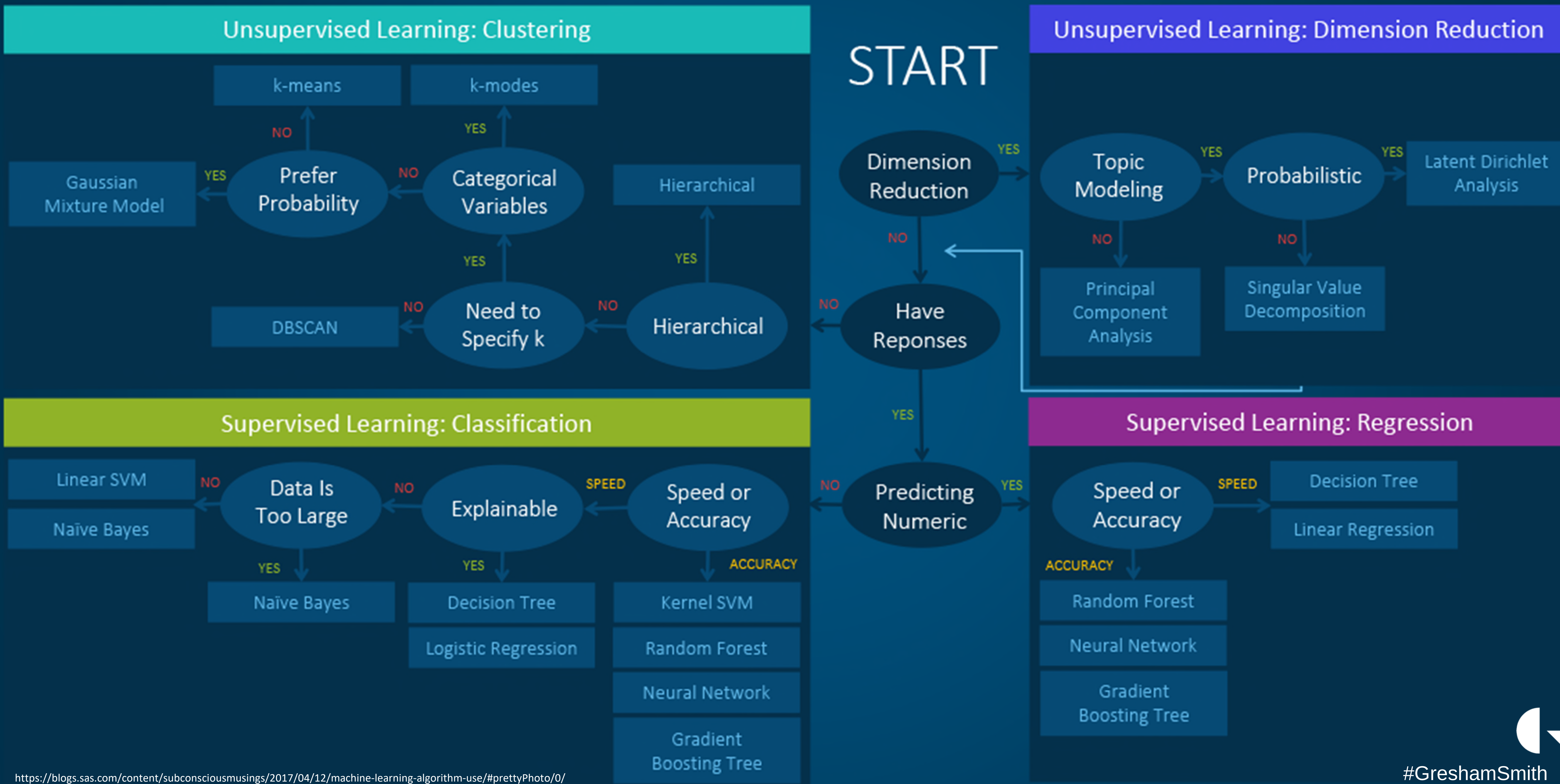


#GreshamSmith

Who Is Using AI Today?

- Nevada, Florida, and Iowa - Deploying Neural Networks for Incident Detection using Video Image Analysis and Traffic Prediction
- Washington DOT - Fuzzy Logic Used for over 20 Years
- California DOT (Caltrans) – Using Fuzzy Logic Metering
- Delaware DOT – Piloted several AI Applications for Traffic Congestion and Incident Prediction
- Metropolitan Transportation Commission of the Bay Area - Light Integration of 511 with Alexa
- Several Arterial Management Agencies - Piloting use of Google Assistant.
- 20+ State DOTs - Active UAS (Drones) Programs, Enhancing with AI in the Future
- Several DOTs - Piloting use of Automated Vehicles for Crash Abatement

Machine Learning Algorithms Cheat Sheet



https://blogs.sas.com/content/subconsciousmusings/2017/04/12/machine-learning-algorithm-use/#prettyPhoto/0/

#GreshamSmith

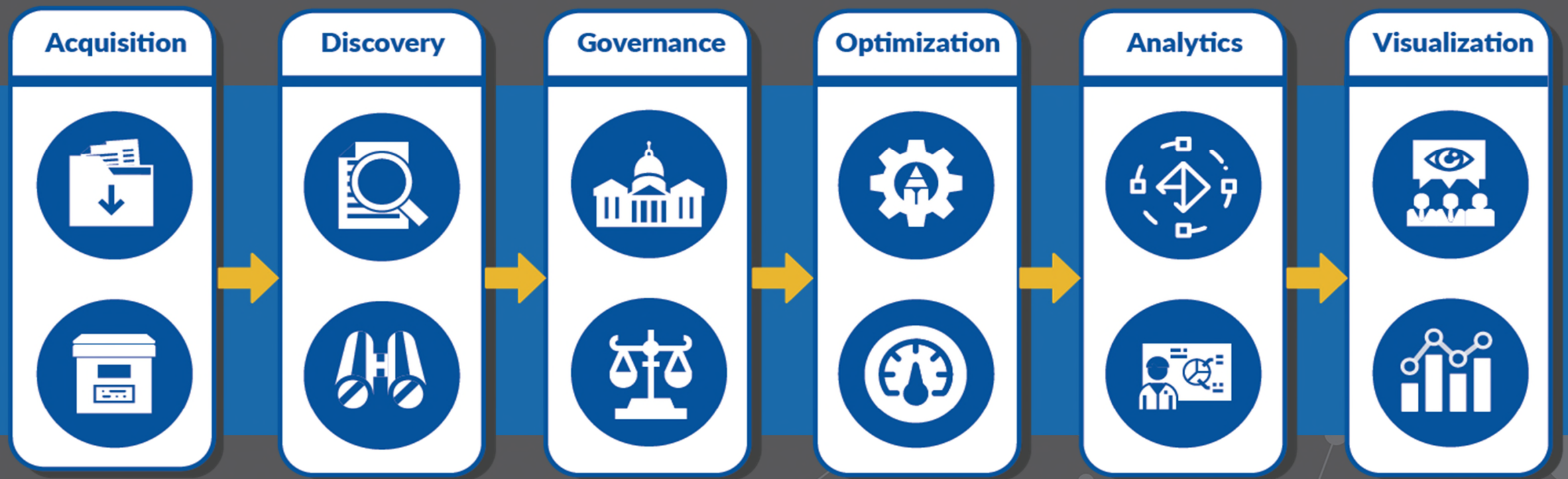
#GreshamSmith

Governing It

Small Steps, Big Results

#GreshamSmith

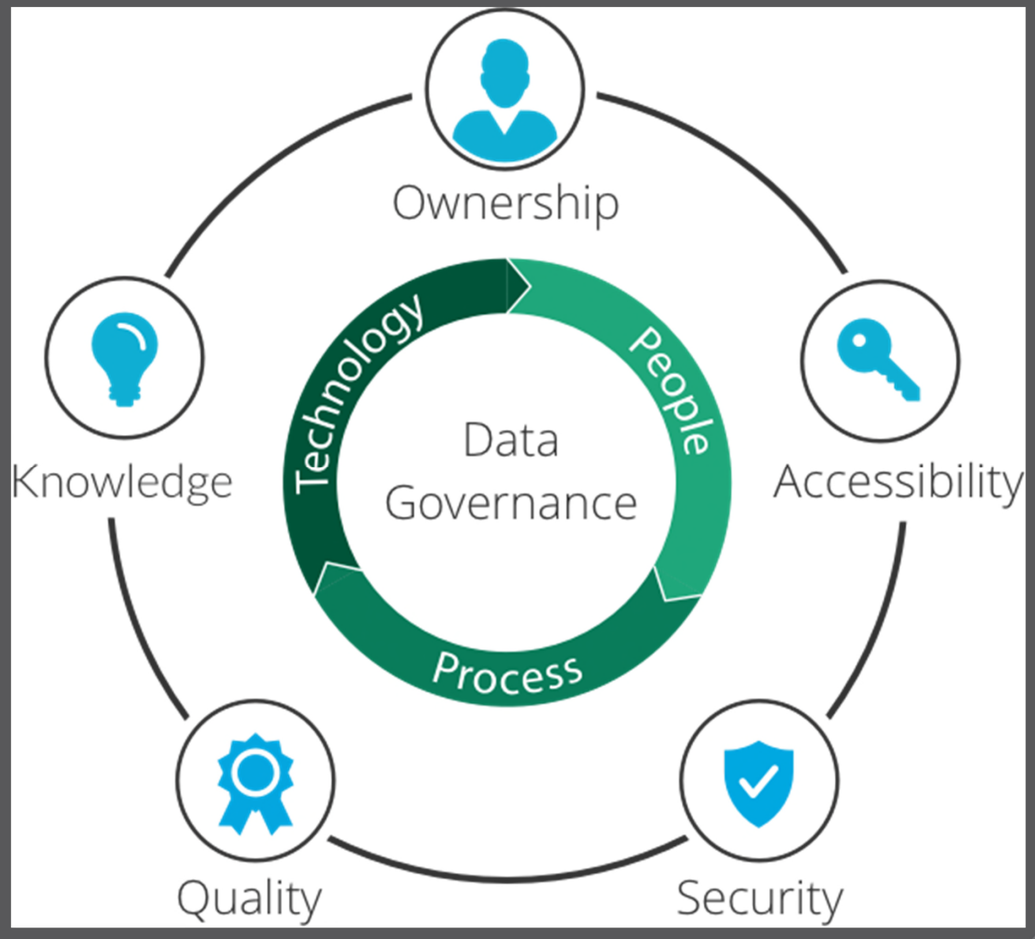
Data Lifecycle



Effective Data Governance - The Difference Between SUCCESS and FAILURE

#GreshamSmith

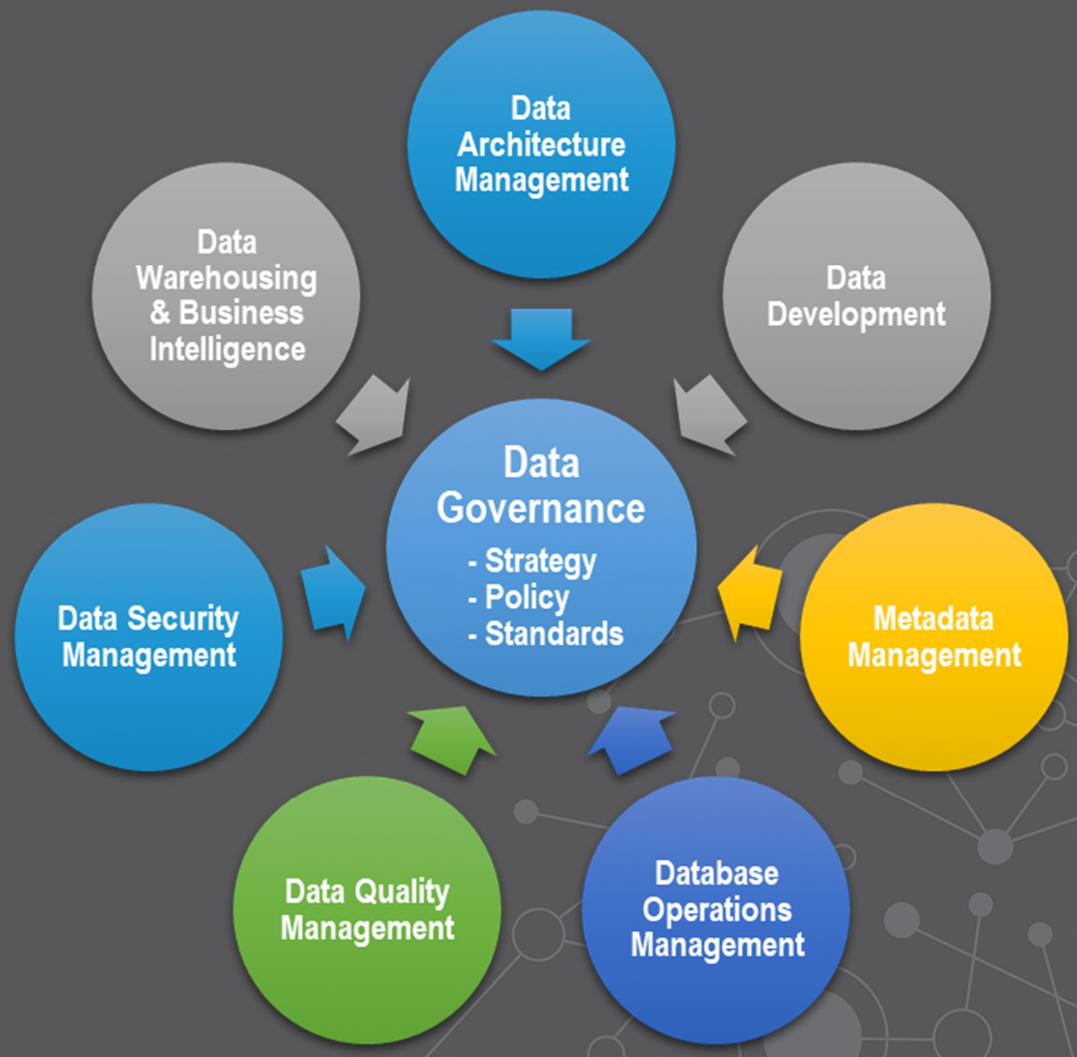
What is Data Governances?



- The Practice of Identifying Important Data Across an Organization, Ensuring it is of High Quality, and Improving its Value to the Business

#GreshamSmith

DATA GOVERNANCE FRAMEWORK

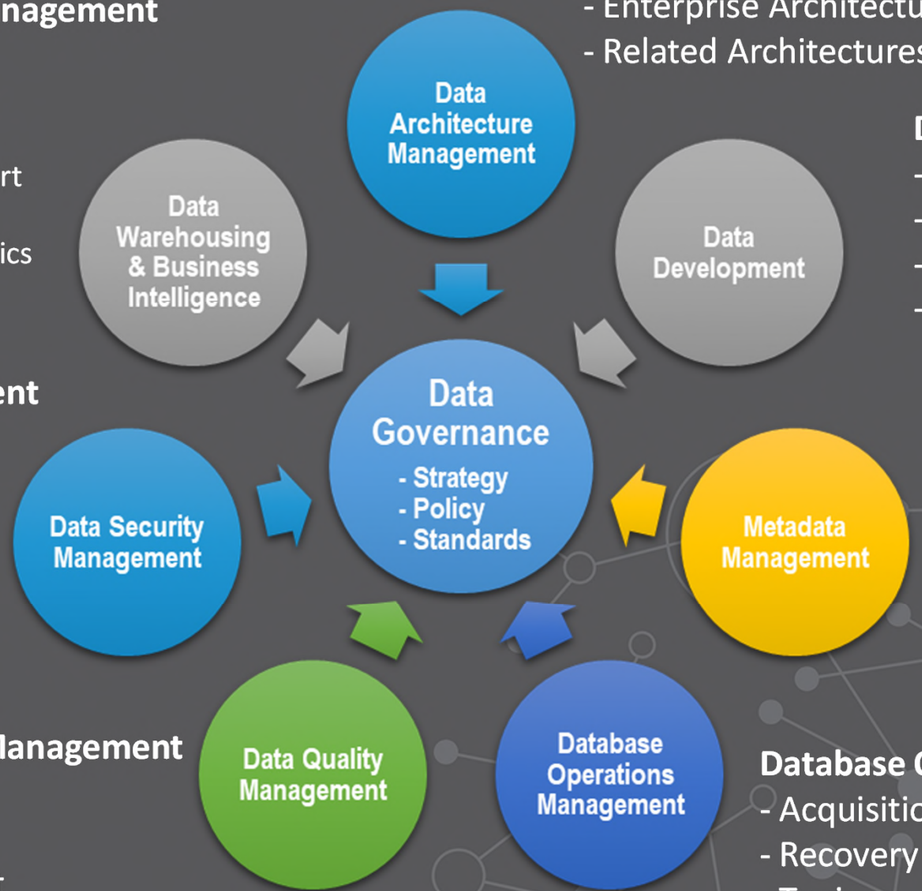


#GreshamSmith

DATA GOVERNANCE FRAMEWORK

Data Architecture Management
- Value Chain Analysis
- Enterprise Architecture
- Related Architectures

Data Warehouse & Business Intelligence Management
- Architecture
- Implementation
- Monitoring
- Training & Support
- Data Mining
- Predictive Analytics



Data Development
- Data Analysis
- Data Modeling
- Database Design
- Implementation

Data Security Management
- Standards
- Classifications
- Administration
- Authentication
- Auditing

Metadata Management
- Architecture
- Integration
- Control
- Delivery

Data Quality Management
- Specification
- Analysis
- Measurement
- Improvement

Database Operations Management
- Acquisition
- Recovery
- Tuning
- Retention

#GreshamSmith

How Do We Do This?

DATA GOVERNANCE – The Lean Way

#GreshamSmith

- Start by Defining the Business Impact
- Acquire Data that is Really Needed
  - Actively Reduce Data Volumes
- Implement Right Data and Analytics Governance
  - Flexible Architecture
    - Greater Consistency, Reusability, and Adaptability
- Automation and Embedded Rules for Ease Data Governance
- Promote Self-Organizing and -Governing Teams
- Ongoing Learning and Refinement Based on Real Data

DATA GOVERNANCE – The Lean Way

#GreshamSmith

Planning and Design

Building

Implementing

Maintaining

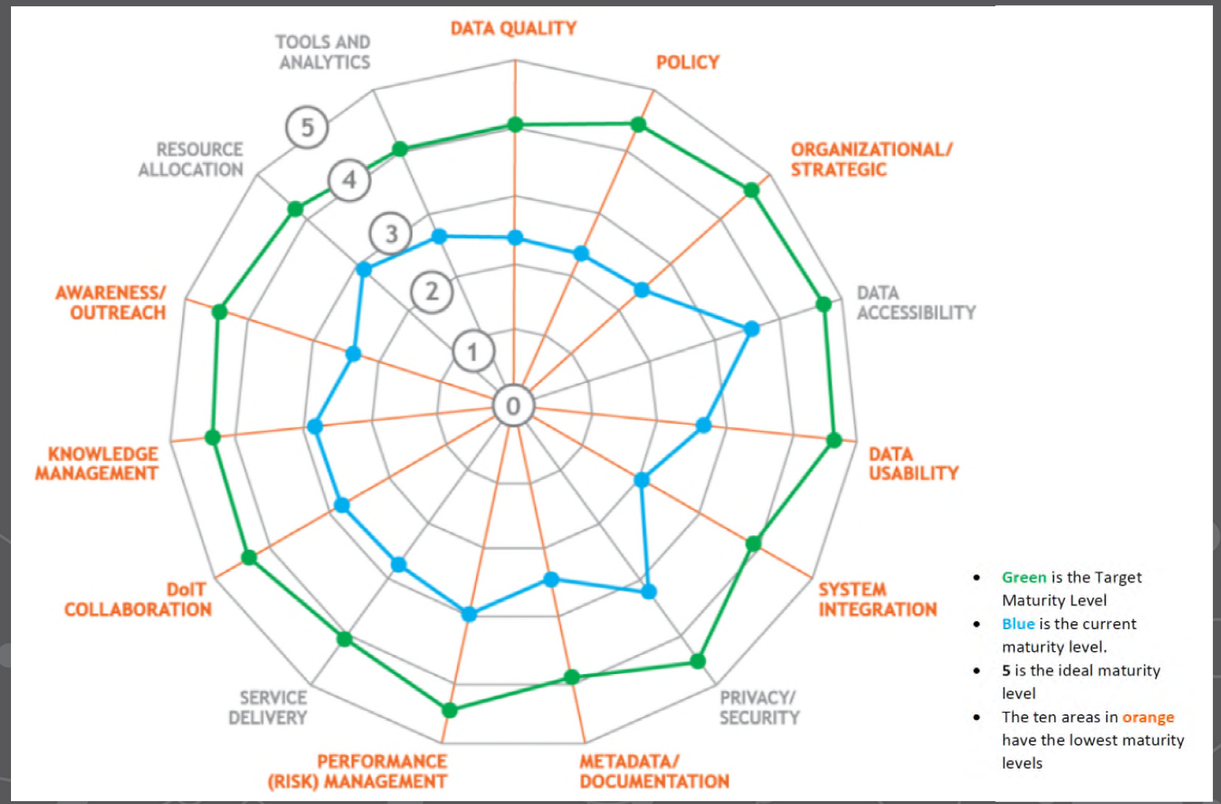
#GreshamSmith

Planning and Design

- Enterprise Data Study
  - What do You Have?
- Data Governance Maturity Assessment
  - People, Policies, Capabilities
- Develop
  - Strategic Plan
  - Implementation Plans
  - Data Standards

Where to Start

- **Understanding Where You Are**
  - Data Capability Maturity
  - Organizational Gaps
  - Explore and Understand Data Sets



Building and Implementing

- Assess Agency Organizational Structure and Skillsets
  - Strategies to Restructure/Augment to Address New Skillsets
- Assess IT's Strategic Plan
- Assess Business Intelligence Capabilities
- Develop and Integrate Data Warehouse
- Standardize Data for New System

#GreshamSmith

Maintaining

- Change Management Plan
  - Resistance Management Strategies
  - Reinforcement Mechanisms
  - Communications Plans
  - Feedback/Metrics
- Executive Level Approved Data Governance Policy
  - Define and Refine Policies and Procedures
  - Effective Rollout Implementation
  - On-going Operations
  - Enforcement
  - Address Data Governance Structure, Data Access, Usage, Quality and Integration

#GreshamSmith

Questions?

Thank you!

#GreshamSmith

Gresham Smith

Genuine Ingenuity

Meredith K Cebelak, PhD, PE

TSM&O Department Leader

Meredith.Cebelak@greshamsmith.com