Finding New Dimensions in Cost Estimating

Pamela S. Wood
Engineering Specialist
The Aerospace Corporation

20 March 2019

Overview

- Removing the Intimidation Factor

- Four Types of Data Analytics

- Examples of Data Analytic Techniques in Cost Estimating

- Barriers to Embracing Data Analytics

- Where Do We End Up?

Removing the Intimidation Factor

Everyone is talking about data analytics, but what does it mean?

- **What is Data Science?**
- Data science is an ***interdisciplinary field*** that uses scientific methods, processes, algorithms and systems to ***extract knowledge and insights*** from data in various forms, both structured and unstructured. [1]
  - *Solid foundations in mathematics, statistics, information science, operations research, and computer science*

- **What is a Data Scientist?**
- Data Scientists ***estimate the unknown*** by asking questions, writing algorithms, and building statistical models [2]
  - *Solid mathematical, computer science and analysis skills*

- **What is Data Analytics?**
- Data analytics is the process of extracting, transforming, loading, modelling, and drawing conclusions from data ***to make decisions*** [3]

- **What is a Data Analyst?**
- Data Analysts ***use data to draw meaningful insights and solve problems*** from well-defined sets of data using an arsenal of different tools to answer tangible questions [2]
  - *Solid mathematical, statistical, and data analytic skills, less programming focused*

Hmmm… this is sounding very familiar…

Removing the Intimidation Factor
Where do Cost Estimators fit?

- Bureau of Labor Statistics: *"**Cost estimators** collect and analyze data in order to estimate the time, money, materials, and labor required to manufacture a product, construct a building, or provide a service. They generally specialize in a particular product or industry."*
  - **Cost Estimate** *– An approximation of the probable cost of a product, program, or project, computed on the basis of available information* [5]

Cost Estimators	Data Analysts / Data Scientists
Collect and Analyze Data for Use in Cost Estimates	Data Analysts & Data Scientists clearly identify the problem they are trying to solve and gather applicable data
Interact with Engineering Teams to Decompose Product, Program or Projects into Time, Money, Materials and Labor Elements	Data Analysts & Data Scientists deconstruct their problem so they can use data to draw meaningful insights to answer questions
Identify Methodologies to Estimate Unknown Costs for Elements	Data Scientists estimate the unknown by writing algorithms and building statistical models from data

Remove the focus on "cost" and all Cost Estimators are already Data Analysts and some are even Data Scientists!

Four Types of Data Analytics
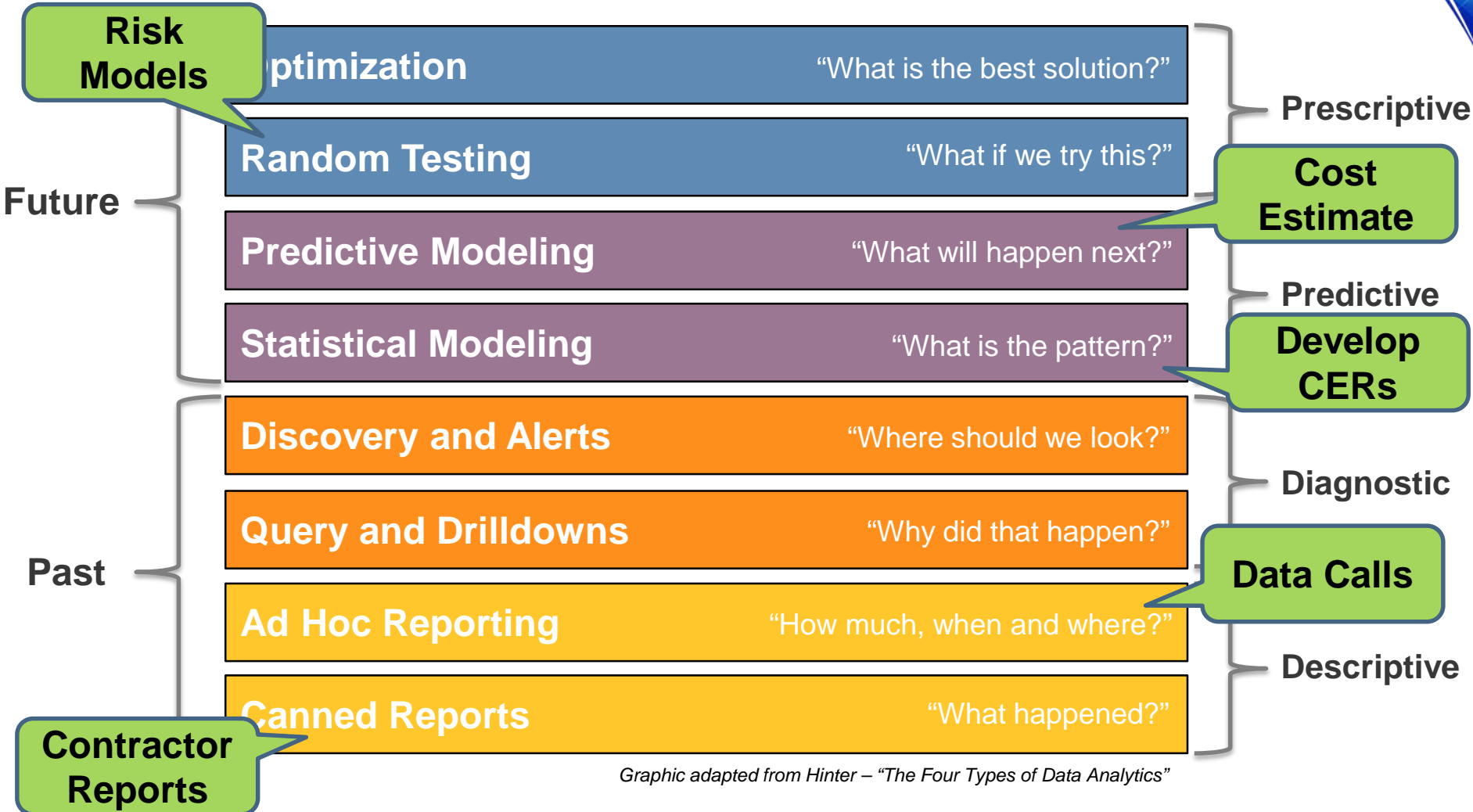*Industry standard types of analytics* [3]

Optimization	"What is the best solution?"
Random Testing	"What if we try this?"
Predictive Modeling	"What will happen next?"
Statistical Modeling	"What is the pattern?"
Discovery and Alerts	"Where should we look?"
Query and Drilldowns	"Why did that happen?"
Ad Hoc Reporting	"How much, when and where?"
Canned Reports	"What happened?"

Future — Optimization / Random Testing / Predictive Modeling / Statistical Modeling

Past — Discovery and Alerts / Query and Drilldowns / Ad Hoc Reporting / Canned Reports

Prescriptive — Optimization / Random Testing

Predictive — Predictive Modeling / Statistical Modeling

Diagnostic — Discovery and Alerts / Query and Drilldowns

Descriptive — Ad Hoc Reporting / Canned Reports

Graphic adapted from Hinter – "The Four Types of Data Analytics"

Different types of analytics are used for different purposes

Four Types of Data Analytics
Analyses Performed in Cost Estimating

Risk Models

Optimization — "What is the best solution?"

Prescriptive

Random Testing — "What if we try this?"

Cost Estimate

Future

Predictive Modeling — "What will happen next?"

Predictive

Statistical Modeling — "What is the pattern?"

Develop CERs

Discovery and Alerts — "Where should we look?"

Diagnostic

Query and Drilldowns — "Why did that happen?"

Past

Data Calls

Ad Hoc Reporting — "How much, when and where?"

Descriptive

Canned Reports — "What happened?"

Contractor Reports

Graphic adapted from Hinter – "The Four Types of Data Analytics"

Standard analyses performed in the cost estimating community fit within these types.

Example 1 – Baby Steps
Using Data Analytics Tools to Perform Descriptive Analysis

Analysis of Federal Employment Data

- Office of Personnel Management (OPM) conducts analyses and develops reports that allow for the assessment of trends and workforce issues that affect OPM and other Federal agencies
  - *Publish the raw data from their database as the* **FedScope Employment Cube**
    - https://www.fedscope.opm.gov/employment.asp

- **Why is this a Data Analytics problem?**
  - FedScope Employment Cube raw dataset contains 25 database structured tables
  - To glean useful information from the dataset, multiple tables need to be combined to perform analysis (i.e., recompile the database and perform a query)
  - The most recent file is 21.6 MB in size
  - Database contains information on 125 different federal agencies
  - The table needed with the pertinent is more than ~3M lines

Size and multiple feeds of data will drive a need for Cost Estimators to perform more complex data normalization and analysis in the future.

Example 1 – Baby Steps

Using Data Analytics Tools to Perform Descriptive Analysis

Save Time: ~5 mins

Write Script to Compile Tables to Provide Specific Information into CSV File saved in folder

Estimated Time: ~30 mins

Run Time: ~2 mins

CSV file provides a smaller dataset able to be handled by Excel

Estimated Time: ~15 mins

The use of a Python script to process and compile large datasets provides efficiency opportunities and reduces size to open access to traditional tools.

Example 2 – Automation
Scalability of Cost Model Excursions

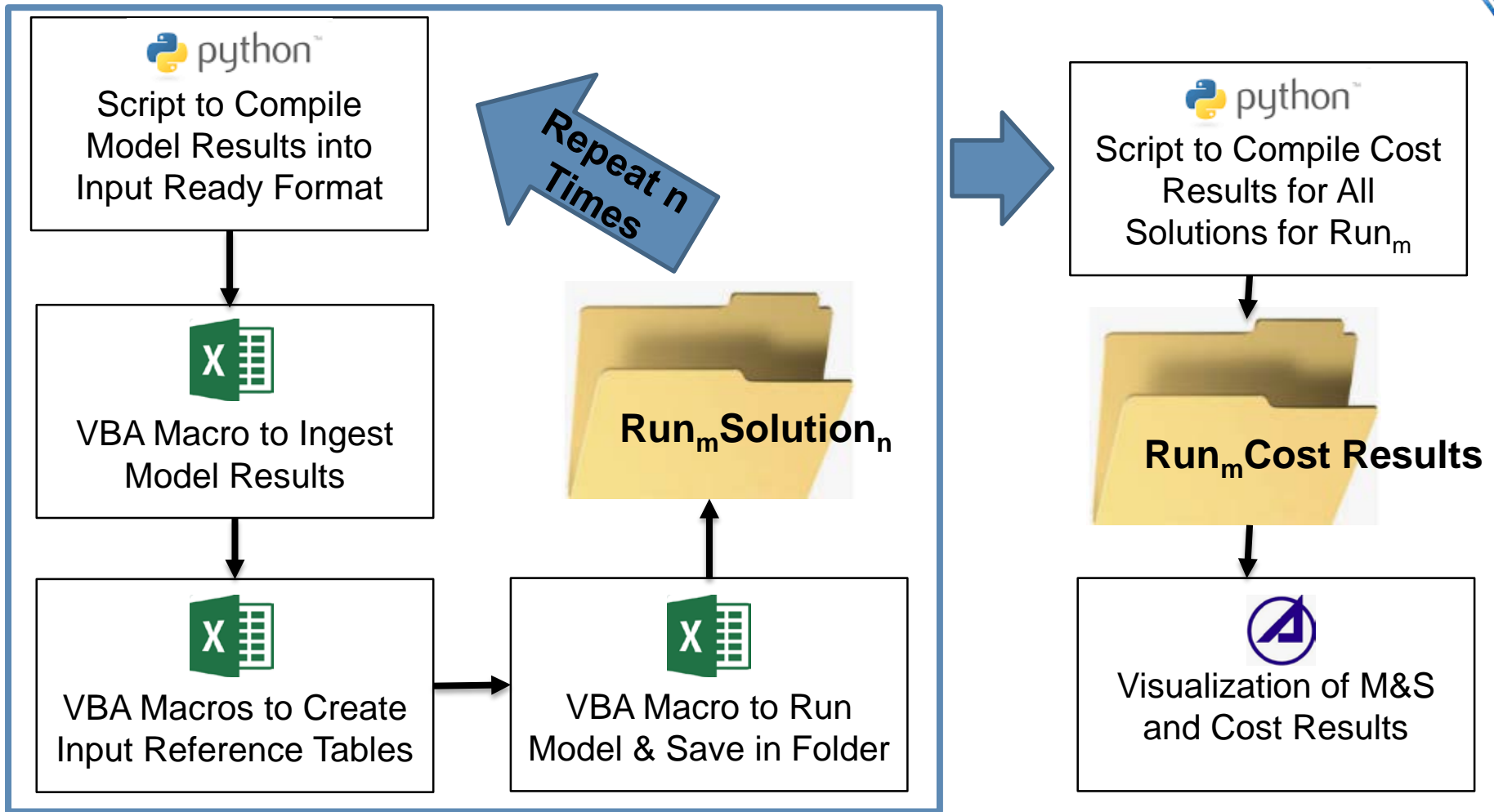## Interfacing with Modeling and Simulation (M&S) Team

- Customer need to utilize modelling techniques to identify optimized performance solutions for a 15 year time period
  - *Optimization runs each produce an unknown number of solutions (lets call this n)*
  - *A cost needs calculated for each solution found by the run*
  - *There is a significant volume of runs (lets call this m)*

- **Why is this a Data Analytics problem?**
  - Significant Volume of Cost Estimates Required
    - Volume of Cost Estimates = Many Feasible Solutions (n) x Volume of Runs (m)
  - Cost estimate results need compiled and returned to M&S team for interactive visualizations
  - Variability of Baseline Architecture to be Costed (e.g. locations, hardware)
  - Project schedule dictated need for quick turn results

Resource constraints, volume and variability of work forced the team to think outside of normal practices and workflows.

Example 2 – Automation

Develop Flexible Cost Model and Automate Repetition

Script to Compile Model Results into Input Ready Format

VBA Macro to Ingest Model Results

VBA Macros to Create Input Reference Tables

VBA Macro to Run Model & Save in Folder

Repeat n Times

$\text{Run}_m\text{Solution}_n$

Script to Compile Cost Results for All Solutions for Run_m

Run_m **Cost Results**

Visualization of M&S and Cost Results

Automation of the Cost Modeling process enables a seamless integration of a multi-disciplinary team working on a highly complex problem.

Example 3 - Collaboration
Same Dataset, Two Organizations

Identification of Redundant Data Collection Efforts

- Two organizations for the same agency collecting data from Quarterly Performance Reports (QPRs) and entering the data into two separate data repositories
  - *Organization #1 – focus on schedule data points*
    - Goal: develop CERs to be used for cost estimation
  - *Organization #2 – focus on cost, performance, schedule and dependencies between programs*
    - Goal: maintain a data repository and dashboard visualization capability to inform stakeholders of status and identify risks for the enterprise

- **Why is this a Data Analytics problem?**
  - *Violates core tenants of data analytics:*
    - Collect once, use many
    - Data becomes stove piped again upon arrival to the agency – defeats purpose of QPR
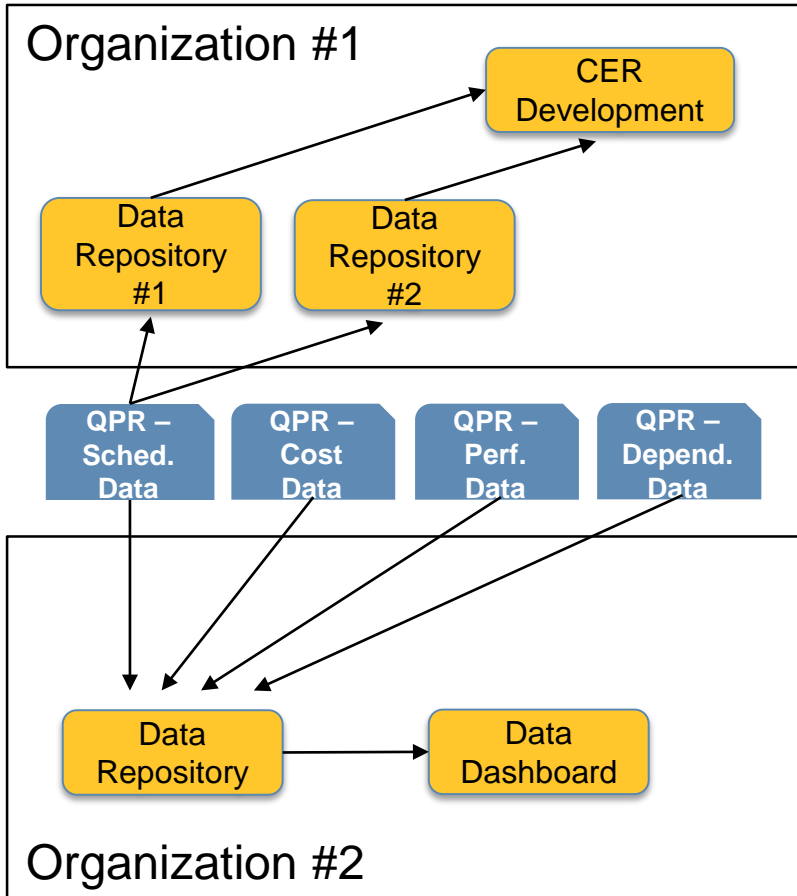   - Inefficient use of resources

Data collection efforts can be optimized to accomplish more than one goal.
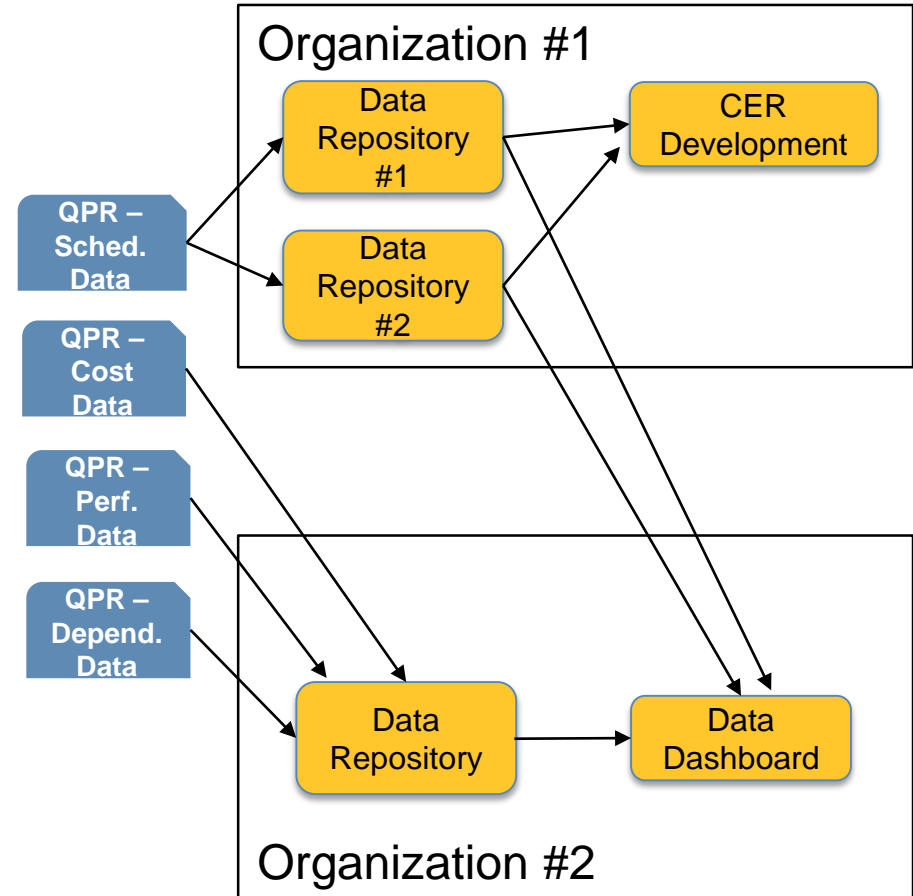
Example 3 - Collaboration

Elimination of Redundancy

Before

Organization #1
- CER Development
- Data Repository #1
- Data Repository #2
- QPR – Sched. Data
- QPR – Cost Data
- QPR – Perf. Data
- QPR – Depend. Data

Organization #2
- Data Repository
- Data Dashboard

After

Organization #1
- Data Repository #1
- Data Repository #2
- CER Development
- QPR – Sched. Data
- QPR – Cost Data
- QPR – Perf. Data
- QPR – Depend. Data

Organization #2
- Data Repository
- Data Dashboard

Eliminating the redundant data collection effort frees up time for deeper analysis.

Barriers to Embracing Data Analytics

The Current Landscape

- **Tools and Capabilities**
  - *Cost Estimating Software*
  - *Home Grown Tools*
  - *Complex Excel Files*

  **Too Rigid?
  Too Flexible?
  Easy to Use?**

- **Organization**
  - *Knowledge Base*
  - *Skill Sets*
  - *Organizational Inertia*

  **Learning
  Organization?**

- **Other Factors**
  - *Data Access and Security*
  - *Policy Restrictions*
  - *Deliverable Timelines*

  **Business
  Restrictions?**

There are a variety of barriers to the community wholly embracing data analytics.

Barriers to Embracing Data Analytics
What Needs to Change?

- To fully embrace data analytics, our community needs to become ***data-driven***[6]
  - *We use data all the time…. Does that make us a data-driven organization?*

- **Data-driven Organizations** [6] **:**
  - *Value the **ongoing sharing of information and collaboration***
  - *Believe all appropriate **data should be accessible** to anyone in the organization*
  - ***Have the tools and skills** to make sense of vast amounts of Structured and Unstructured data to inform decisions and make predictions*
  - ***Make data collection a primary activity** for every part of the organization and reward and measure the behavior*
  - ***Provide real-time insights** to identify where course corrections are needed and new opportunities present themselves*
  - ***Senior leaders recognize and support** the data-driven culture*
  - *Use data to **constantly diagnose problems** in the system and processes*
  - ***Value transparency** and upward communication*

Using and analyzing data does <u>not</u> mean your organization has a data-driven culture.

Barriers to Embracing Data Analytics
How do we get there?

Steps toward evolving into a data-driven enterprise[7]:

1. **Develop an integrated data warehouse**
   - *Master data repository where data is captured only once*
   - *Enables the consistent use of data across the organization*
   - *Facilitates tighter security controls*

2. **Make data readily available**
   - *Provide transparency to the data*
   - *Create flexible reporting capabilities*

3. **Understand limitations**
   - *Technical limitations to data gathering*
   - *Anticipate down the road issues associated with data traceability and level of fidelity*

4. **Harness tools**
   - *Provide meaningful insights*
   - *Predict and manage risk*
   - *Reveal new areas of opportunity*

5. **Conduct Business Intelligence and Data Analytics (BIDA) health-checks**
   - *Ensure people, processes and things aligned with data strategy and roadmap*
   - *Once or twice a year, including seniors, to assess the vision*

Small steps make huge differences. Jumping the gun can result in failure.

Where Do We End Up?

What does a Cost Estimating world look like with Data Analytics?

- Build from a knowledge management foundation
  - *Consists of the initiatives, processes, strategies, and systems that sustain and enhance the storage, assessment, sharing, refinement, and creation of knowledge*[1]
  - *Data Collection and Normalization still remains a primary function but the 80/20 rule begins to shrink*

- Evolve into a learning organization
  - *A learning organization is a company that facilitates the learning of its members and continuously transforms itself*[1]
  - *Documentation of institutional knowledge*

- Improved organizational efficiency
  - *Monitor CERs overtime for adjustments and continued validity*
  - *Models can become plug and play by tapping different data libraries and/or vetted estimating methodologies – Similar to Cost Estimating Software but on a more tailored and larger scale*
  - *Repetitive tasks can be eliminated*

Data Analytic techniques open unexplored new dimensions.

Questions?

References

1. Wikipedia. https://en.wikipedia.org/wiki/
2. Lebowitz, Juliana. *Data Analytics vs Data Science: The Breakdown.* Northeastern University. 2018 July 16. https://www.northeastern.edu/levelblog/2018/07/16/data-analytics-vs-data-science/
3. Hitner, Nicole. *The Four Types of Data Analytics*. Datafloq. 2018 November 03. https://datafloq.com/read/the-four-types-of-data-analytics/3903
4. Occupational Outlook Handbook, Bureau of Labor Statistics. https://www.bls.gov/ooh/business-and-financial/cost-estimators.htm#tab-1
5. Business Dictionary. http://www.businessdictionary.com/definition/cost-estimate.html
6. Gleeson, Brent. *The Benefits of Leading Data-Driven Organizational Change*. Forbes. 2017 September 28. https://www.forbes.com/sites/brentgleeson/2017/09/28/the-benefits-of-leading-data-driven-organizational-change/#718732c95e09
7. Terrill, Stephanie, Sanjiv Purba. *Data-driven Business Transformation*. KPMG. 2015.