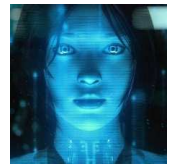
The Ten V's of Big Data

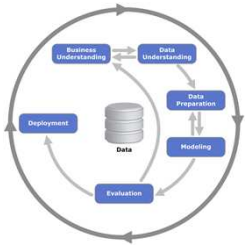
(Understanding Data Risks)





Richard Frederick, PMP (rfrederick.pmp@gmail.com; 214-755-7035 talk or text; www.linkedin.com/in/rfrederick)



CBDA CERTIFIED IIBA

ISO International Organization for Standardization

ANSI American National Standards Institute

sas

OMG OBJECT MANAGEMENT GROUP

INCOSE International Council on Systems Engineering

IEEE

PMI Project Management Institute

Scaled Agile Framework

IIBA International Institute of Business Analysis

UNIFIED MODELING LANGUAGE

W3COMPUTING A DEVELOPER REFERENCE WEBSITE

Requirements Development Management (RDM) for

Business Data Analytics

(The Language of Data)

Richard Frederick, PMP (rfrederick.pmp@gmail.com; 214-755-7035 talk or text; www.linkedin.com/in/rfrederick)

Requirements Agenda

Business Data Analytics "The Language of Data"

01-FOUNDATION

- **LAB-Introductions (Name, Job Title, Objectives)**
- Dashboards and Robots (Data Mining & Machine Learning)
- Requirements and Testing (Four Quadrants)
- What are Structured Language Requirements? (Structured English and Structured Query Language)
- Why Should You Care? (Primary Source of Project Problems)
- How Do They Work? (Discreet Intellectual Property Inventory)
- Types of Requirements (Product, Project, DATA)
- Natural Language Processing (Morphology, Semantics, Syntax and Linguistics)
- OMG-SBVR (Semantics of Business Vocabulary & Rules)
- IEEE-EARS (Easy Approach to Requirements Syntax)
- INCOSE (Rules for Writing Requirements) & QVscribe
- Waterfall and Agile (Assembly Methods)

02-ELICIT

- **LAB-Vision/Scope (Seek to Understand)**

Elicitation Techniques:

- Document Analysis (Low Hanging Fruit)
- Interface Analysis (Navigation & Functionality)
- Benchmarking (Actual Data)
- Brainstorming (Every Idea is a Good Idea until it becomes a Bad Idea)
- Prototyping (Minimum Viable Product)
- Reverse Engineering (Begin with the End in Mind)
- Interview (Thinking Questions)
- Workshop (Group Interviews)
- Observation (What do you See?)
- Survey Questionnaire (Paper equals proof)

03-ANALYZE

- What are Models? (Pictures of Language)
- **LAB-The Language of Modeling (GIVEN pre WHEN process THEN output-result)**
- Types of Models (Context-Structure, Usage, Data Behavior, Process Flow)
- Context-Structure (Vision, Roadmap, Scope WBS)
- Usage (EPIC, UseCase, UserStory, Feature)
- Data Behavior (ERD, JOIN-Denormalization, Star Schema, Dimensional OLAP, Dashboard, Intelligence)
- Data Behavior (Data Dictionary, DataFlow, Data Structure Instance, Data Element Attribute, Data Store)
- Data Behavior (Process Logic, Business Rules)
- Process Flow (Swimlane)

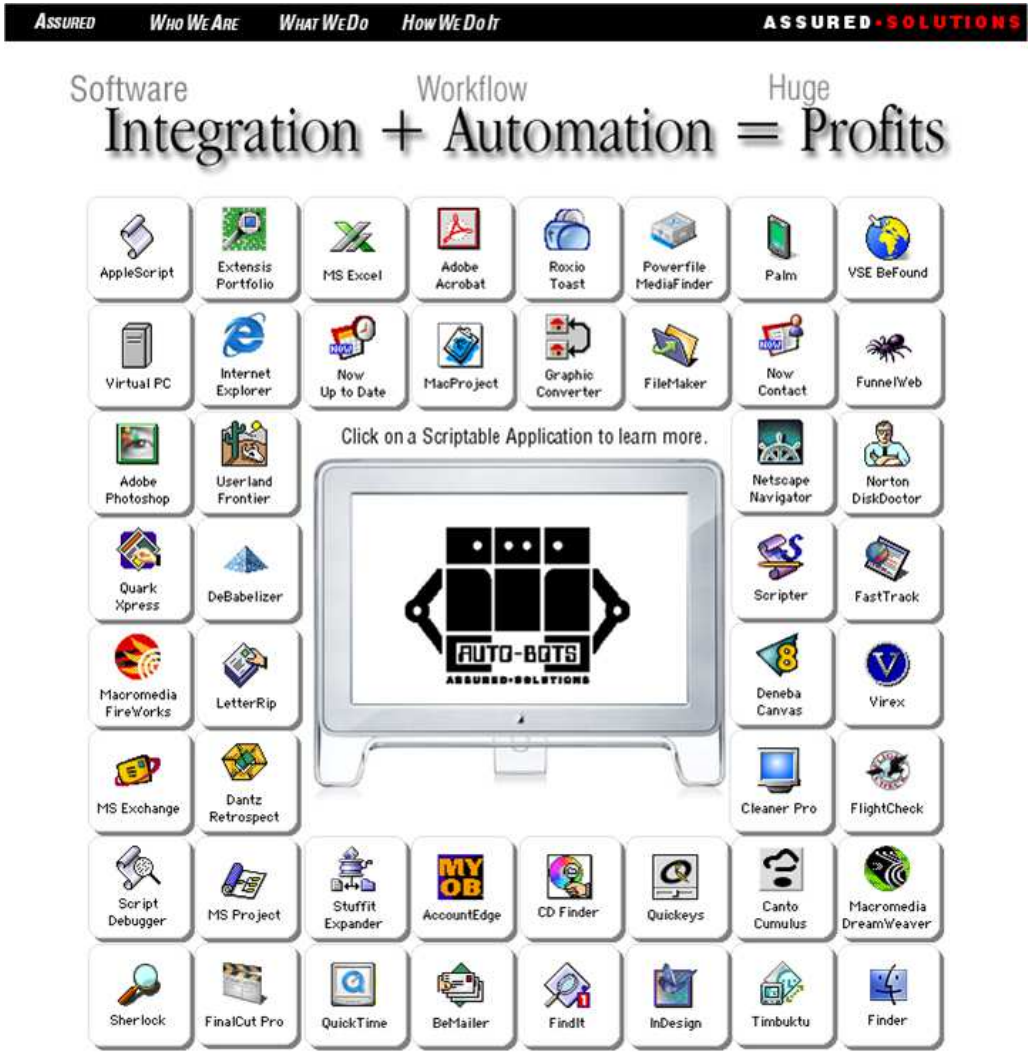
04-DOCUMENT

- Categorization, Organization, Documentation, Integration, Automation
- Making Documents Easy to Read (Fonts & Navigation)
- Document Types (BRD, TRD)
- **LAB-Business Requirement Document (Concept of Operation)**
- Technical Requirement Document (System Specification)

05-VALIDATE

- Validation thru Triangulation (Prep Drills)
- Traceability (Project Unique Identifier)
- Requirements Baseline (ROM Estimate, Planning Estimate, Definitive Estimate)
- **LAB-Estimating Story Points (Complexity and Risk)**
- Lessons Learned (Course Wrap-Up)

Assured Solutions

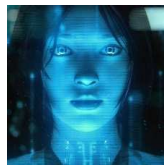


Overview "Seek to Understand"

- What is Big Data?
  Everything, Quantified and Tracked
- Advantages of Big Data
- Ten V's of Big Data
- Understanding Data Risk





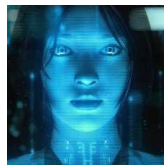



What is Big Data?

- Everything (**Every aspect of life**, work, consumerism, entertainment, and play is now recognized as a source of digital information)
- Quantified (Storing "everything" somewhere, mostly in digital form, often as numbers.)
- Tracked (Quantifying and measuring "everything" continuously)





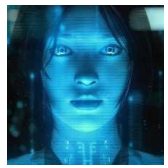



What is Big Data?

	"Big" Data	"Small" Data
VOLUME:	More than a human can comprehend	A lot, but a human brain can handle it
VELOCITY:	Gigantic waves of data every millisecond	Steady stream throughout the data
VARIETY:	Unknown file types	Expected set of file types
VERACITY:	Almost no idea of how data might come in or in what format	Know where changes might happen
STRUCTURE:	Non-Relational	Relational
HARDWARE:	Data Center	File Server
DATABASE:	NoSQL	SQL
PROCESSING:	Spark, Hive, or Pig (on TOP of Hadoop)	SQL
QUERY LANGUAGE:	Python, Java, R, Sequel	SQL
ANALYSIS AREAS:	Clusters (Data Scientist) Data Marts (Analytics)	Data Marts (Analytics)
OPTIMIZATION:	Machine Learning	Manual, human powered
PEOPLE:	Data Scientists, Data Engineers, DBAs, Analysts	Data Engineers, DBAs, Analysts
NOMENCLATURE:	Data Lake	Database, Data Warehouse, Data Mart





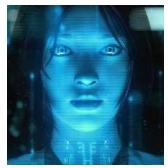
Advantages of Big Data

- Smarter Decisions
- Better Products
- Deeper Insights
- Greater Knowledge
- Optimal Solutions
- Customer-Centric Products
- Increased Customer Loyalty

- More Automated Processes,
- More accurate Predictive and Prescriptive Analytics
- Better models of future behaviors and outcomes in Business, Government, Security, Science, Healthcare, Education, and more.





Ten V’s of Big Data

V's of Big Data

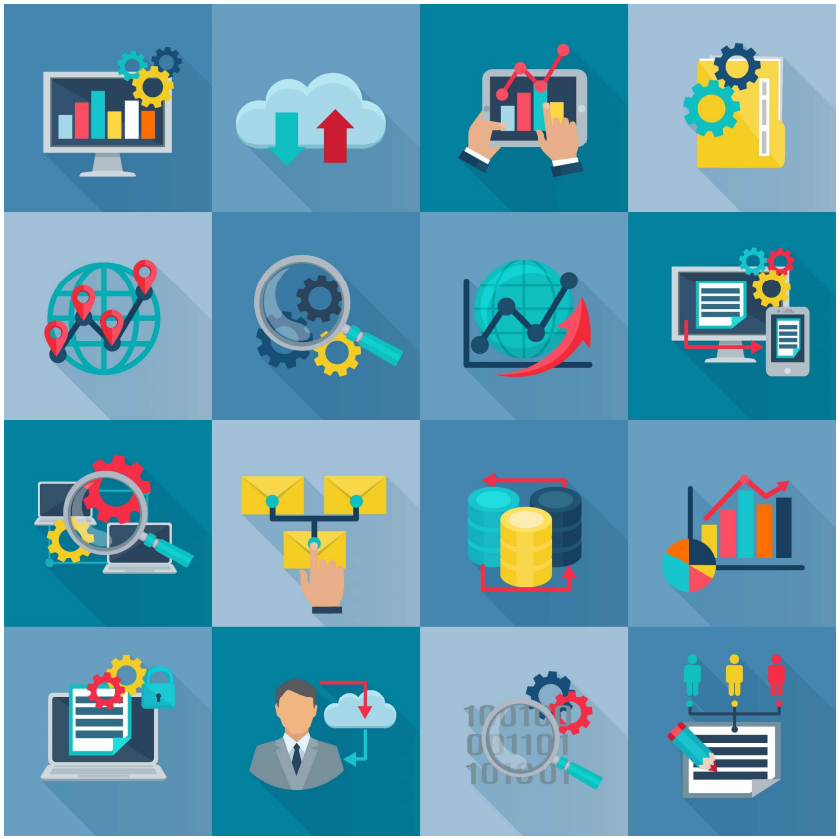
1 Value: *How Useful is the Data?* (Hypotheticals, Statistical, Events, Correlations)

2 Volume: *Size of the Data?* (Distributed, Tables, Records, Exabytes)

3 Velocity: *Data Generation Speed?* (Real-time, Processes, Streams)

4 Variety: *Data Types?* (Unstructured, Structured, Multi-factor, Probabilistic, Linked, Dynamic)

5 Veracity: *Data Accuracy?* (Accountable, Trustworthy, Authentic)



V's of Big Data

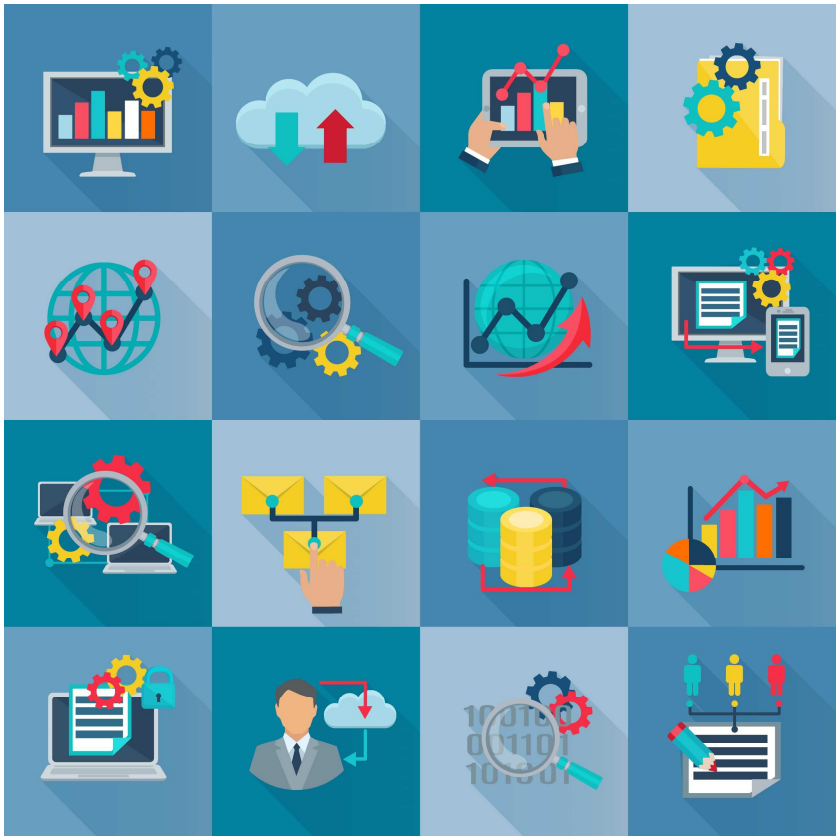
6 Variability: *Evolving Behavior in Data Source?* (Changing Data, Changing Model)

7 Validity: *Data Quality, Governance & Master Data Management?*(Uniformity, Accuracy, Stewardship, Semantic Consistency)

8 Venue: *Multiple Heterogeneous Platforms?* (ERP, CRM, Facebook, LinkedIn, Twitter)

9 Vocabulary: *Data Structure Description?* (Data Dictionary, Data Flow, Data Structure Instance, Data Element Attribute, Data Process, Data Store)

10 Vagueness: *Terminology Confusion?* (Semantic Community, Taxonomy, Business Rules)



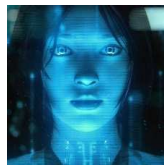
Understanding Data Risks (Testing)

- Data Integration Complexity
- Data Security, Privacy, Governance
- Lack of Meta Data, & Schema for Big Data

- Poor Data Quality:
  - Noisy Data (Corrupted)
  - Dirty Data (Inaccurate)
  - Misplaced Data Values
  - Inexact Values
  - Insufficient Data Size
  - Poor Representation In Data Sample
  - Redundant Data





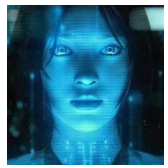



Summary "Train the Trainer"

- What is Big Data?
  Everything, Quantified and Tracked
- Advantages of Big Data
- Ten V's of Big Data
- Understanding Data Risk





Let's stay in contact:

Richard Frederick, PMP

214-755-7035 (text or talk)

Rfrederick.pmp@gmail.com

www.meetup.com/tampa-bay-IIBA/

www.linkedin.com/in/rfrederick

meetings.hubspot.com/rfrederick-pmp