Data Literacy Terms & Concepts

By Dr. David P. Marco, Fellow IIM, CDP, CBIP

President

EWSolutions

EWSolutions

EWSolutions is a Chicago-headquartered strategic partner and full life-cycle systems integrator providing both **award winning** strategic consulting and **full-service implementation services**. This combination affords our client partners a full range of services for any size enterprise information management, metadata management, data governance and data warehouse/business intelligence initiative. Our notable client partner projects have been featured in the Chicago Tribune, Federal Computer Weekly, Journal of the American Medical Informatics Association (JAMIA), Crain's Chicago Business, FBI, The Doings and won the 2004 Intelligent Enterprise's RealWare award, 2007 Excellence in Information Integrity Award nomination, DM Review's 2005 World Class Solutions award, 2016 CIO Review 20 Most Promising Enterprise Architecture providers and 2018 Innovation award.

FBI
Meritorious Service Award

2007 Excellence in Information Integrity Award Nomination

Best Business Intelligence Application Information Integration Client: Department of Defense

Excellence and Thought Leadership

2018: Most Innovative in Data Management Solutions

World Class Solutions Award Data Management

For more information on our Strategic Consulting Services, Implementation Services, or World-Class Training, email us at info@EWSolutions.com

Over 155 Successful Client Partner Implementations
Contact us at INFO@EWSolutions.com to become our next success

GSA Schedule

AFLAC
AgVantis
AON (Hewitt Associates)
Arizona Supreme Court
Bank of America (Countrywide)
Bank of Montreal
BankUnited
Basic American Foods
Becton, Dickinson and Company
Blue Cross Blue Shield Association
Blue Cross Blue Shield companies
Booz Allen Hamilton
Branch Banking & Trust (BB&T)
British Petroleum (BP)
California DMV
California State Fund
Canadian Institute for Health Information (CIHI)
Canadian National Railway
Capella University
Centers for Medicare & Medicaid Services (CMS)
Cigna
College Board
Comcast
Corning Cable Systems
Contech Systems, Inc.
CostCo
D.A. Davidson
Defense Logistics Agency (DLA)

Delta Dental
Department of Defense (DoD)
Department of Homeland Security (DHS)
Department of the Interior
Discover
Driehaus Capital Management
Eli Lilly and Company
Environment Protection Agency (EPA)
Erwin
Fannie Mae
Farmers Insurance Group
Federal Aviation Administration
Federal Bureau of Investigation (FBI)
FedEx
FEMA
Fidelity Information Services
FICO (Fair Issac)
Ford Motor Company
GlaxoSmithKline
Harbor Funds
Harris Bank (BMO)
The Hartford
Harvard Pilgrim HealthCare
Health Care Services Corporation
Henny Penny
HP (Hewlett-Packard)
IDMA (Insurance Data Mgt. Institute)
Information Resources Inc.
International Paper
IQVIA (Quintiles)
Janus Mutual Funds

Johnson Controls
Key Bank
LiquidNet
Loyola Medical Center
Manulife Financial (John Hancock Fin.)
Mayo Clinic
McDonalds
Micron
Microsoft
MoneyGram
NASA
National City Bank (PNC Financial)
Nationwide
Neighborhood Health Plan
NORC
Physicians Mutual Insurance
Pillsbury
Quartz Benefits
Rowan College of South Jersey (RCSJ)
Sallie Mae
Saudi Arabia Minister of the Interior
Saudi Telecom Company (STC)
Schneider National
Secretary of Defense/Logistics
Singapore Defense Science & Technology Agency (DSTA)
Social Security Administration
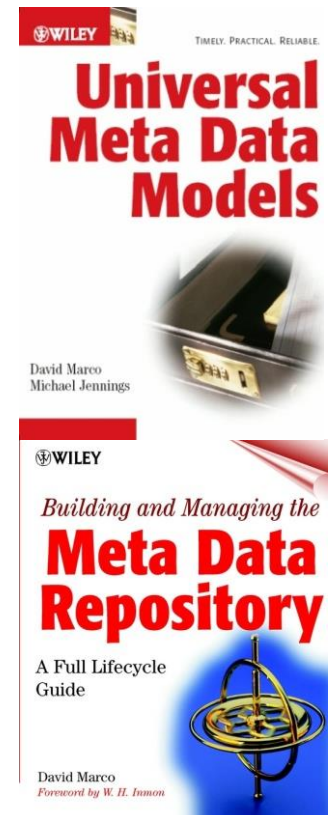South Orange County Community College
Spherion

Standard Bank of South Africa
SunBelt Rentals
SunTrust Bank
Surescripts
Target Corporation
Texas CPA
The Regence Group
Thomson Multimedia (RCA)
Thrivent Financial
United Healthcare (UHC)
United Health Group
United Nations (ICAO)
United States Air Force
United States Federal Courts
United States Department of State
United States Fish and Wildlife Service
United States Navy
United States Navy (BUPERS)
United States Transportation Command
University of Michigan
University of Wisconsin Health
USAA
US Cellular
Vera Bradley
Waste Management
Wells Fargo
Wisconsin Department of Transportation
Workers' Compensation Insurance Rating Bureau of California (WCIRB)
Zurich Cantonal Bank

For more information on our Strategic Consulting Services, Implementation Services, or World-Class Training email us at **Info@EWSolutions.com**

Best known as the world's foremost authority on metadata management and the father of the Managed Metadata Environment, he is an internationally recognized expert in the fields of data governance, digital transformation, data literacy, big data, data warehousing, master data management and data management. In 2004 Dr. David P. Marco was named the **"Melvil Dewey of Metadata"** by **Crain's Chicago Business** as he was selected to their very prestigious **"Top 40 Under 40"** list. Dr. Marco has authored several books including the widely acclaimed **"Universal Metadata Models"** (Wiley, 2004) and the classic **"Building and Managing the Metadata Repository: A Full Life-Cycle Guide"** (Wiley, 2000).

- ❏ 2020 received a lifetime professional **Fellowship from the Institute of Information Management** (their highest honor)
- ❏ President of Data Management University (**DataManagementU.com**)
- ❏ Author of several best-selling information technology books, including the top 2 sellers in metadata management history
- ❏ **2016 Data Management Channel Expert** for Business Analytics Collaborative
- ❏ **2008 DAMA Data Management Hall of Fame (**Professional Achievement Award**)**
- ❏ **2007 DePaul University named** him one of their **"Top 14 Alumni Under 40"**
- ❏ Selected to the prestigious **2004 Crain's Chicago Business "Top 40 Under 40"**
- ❏ Worked with over 90 client partners to successfully build their data management & governance programs
- ❏ Presented hundreds of keynotes/seminars across four continents
- ❏ Published hundreds of IT articles some of which were translated into Mandarin, Russian, Italian, Portuguese and others
- ❏ Taught at the **University of Chicago** and **DePaul University**
- ❏ Holds PhD, Fellow IIM, CDP, CCP and CBIP certifications

Email: DMarco@EWSolutions.com

WILEY
TIMELY. PRACTICAL. RELIABLE.
Universal Meta Data Models
David Marco
Michael Jennings

WILEY
Building and Managing the
Meta Data Repository
A Full Lifecycle Guide
David Marco
Foreword by W. H. Inmon

DATAMANAGEMENT EDUCATION.COM

DATA MANAGEMENT UNIVERSITY
YOUR PORTAL FOR DATA MANAGEMENT KNOWLEDGE SINCE 1999

- ❑ Data Literacy for the Individual
- ❑ Data Literate Thinking
  - ➢ Critical Thinking
  - ➢ Scientific Thinking
  - ➢ Statistical Thinking
  - ➢ Visual Thinking
  - ➢ Skepticism
  - ➢ Ethical Thinking
- ❑ Data Literacy for the Organization

- ❑ Basic Data Literacy Terms and Concepts
  - ➢ Inference
  - ➢ Null Hypothesis
  - ➢ Statistical Significance
  - ➢ Data Wrangling/Data Munging
  - ➢ Signal
  - ➢ Noise
  - ➢ Data Democratization
  - ➢ Frequency
  - ➢ Correlations
- ❑ Explanatory Analysis vs. Exploratory Analysis
- ❑ Standard Deviation
- ❑ Normal Distribution

☐ ***Data Literacy:*** *is the ability to create, manage, read, work with, and analyze data to ensure and maximize the data's accuracy, trust, and value to the organization*

Data Literacy for the Individual

❑ A data literate person can:

➢ Understand data – what it represents, and what it does not (what is missing, unclear, wrong, etc.)

➢ Find and obtain data – where to obtain the right data for the stated purpose

➢ Read, interpret and evaluate data – what is the data indicating, and the reasons for these results

➢ Manage data – knowing the rules, standards, processes for effectively managing and using data

➢ Create insight, using data and data sets – presenting the results in appropriate formats to the right audiences to enable operations and decisions

EWSolutions™

DATAMANAGEMENT EDUCATION.COM

Ethical Thinking – u*nderstanding the potential for good or for harm of any actions or conclusions from data; to actively prevent harm*

Critical Thinking – *being open to different assumptions or conclusions; to be willing to adopt new ideas and perspectives on a subject*
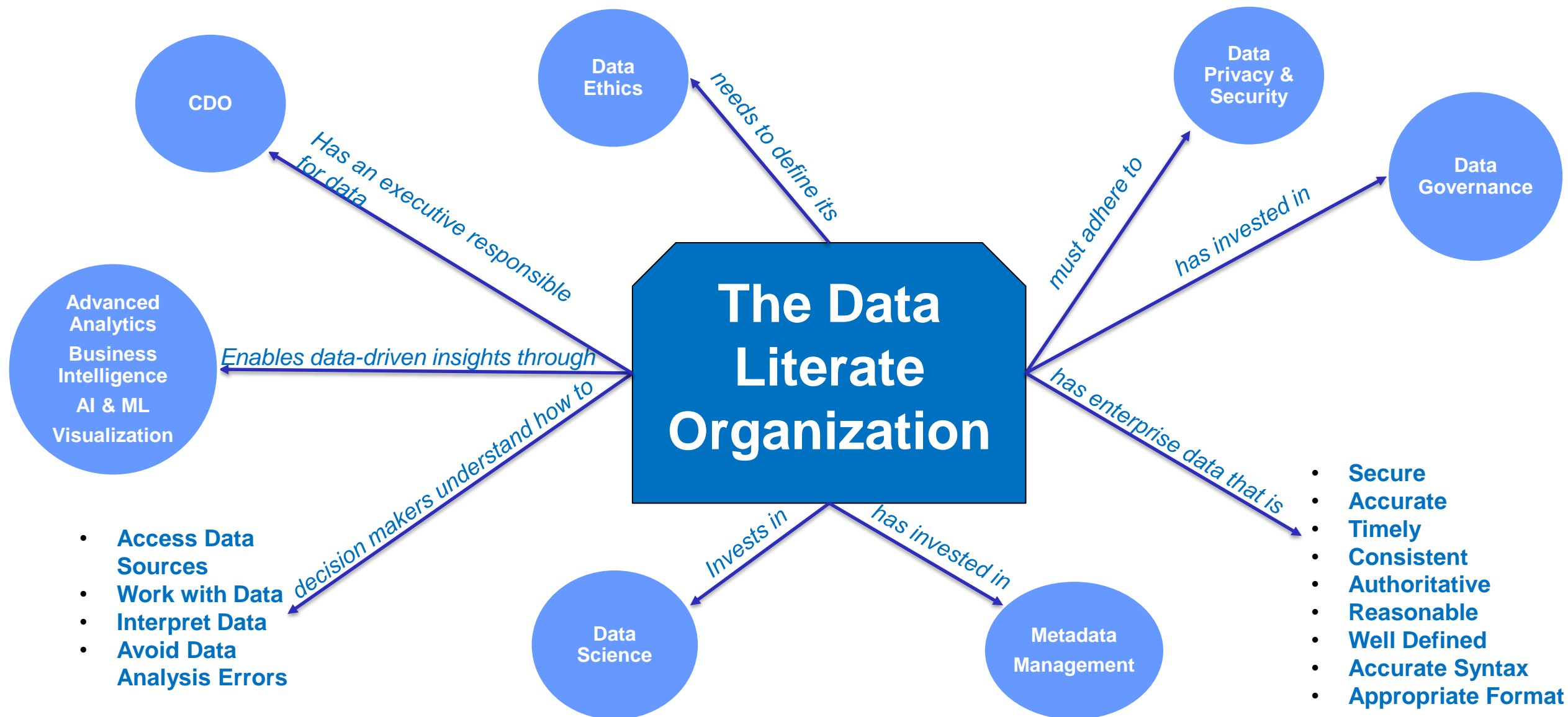
Scientific Thinking – k*nowledge of the scientific method, scientific principles, and the ability to apply a scientific approach to formulate and test hypotheses*

Skepticism – Willingness to question the data, to go beneath the surface, to understand the context in which data is created and presented

Data Literate Thinking

Visual Thinking – *the ability to understand and interpret information conveyed visually, through graphs, charts, and other means*

Statistical Thinking – k*nowledge of how statistics work and how to apply them and to apply quantitative reasoning to problems*

Data Literacy for the Organization

The Data Literate Organization

EWSolutions™

DATAMANAGEMENT EDUCATION.COM

Data Ethics

CDO

Data Privacy & Security

Data Governance

Advanced Analytics

Business Intelligence

AI & ML

Visualization

needs to define its

Has an executive responsible for data

Enables data-driven insights through

must adhere to

has invested in

decision makers understand how to

has enterprise data that is

Invests in

has invested in

- **Access Data Sources**
- **Work with Data**
- **Interpret Data**
- **Avoid Data Analysis Errors**

Data Science

Metadata Management

- **Secure**
- **Accurate**
- **Timely**
- **Consistent**
- **Authoritative**
- **Reasonable**
- **Well Defined**
- **Accurate Syntax**
- **Appropriate Format**

Over 155 Successful Client Partner Implementations
Contact us at INFO@EWSolutions.com to become our next success

© 1997 to present, Enterprise Warehousing Solutions, Inc. (EWSolutions) – 11

Basic Data Literacy Terms and Concepts

- **Inference:** is the process of drawing conclusions about a population based on a sample of the data. This occurs because, in most instances, it is not practical to obtain all the measurements in a given population.
- **Null Hypothesis:** assumes there is no statistical importance between the two variables in the hypothesis.
- **Alternative Hypothesis:** supports a statistically significant result. The opposite of a null hypothesis. Rejects the null hypothesis.
- **Statistical Significance:** a number that expresses the probability that the result of a given experiment (analysis) or study could have occurred purely by chance.
  - Margin of error (Results of the survey are accurate to 7%)
  - Confidence level (If a poll is repeated, there is an 85% probability that the conclusions would be validated).
- **Data Wrangling/Data Munging:** is the process of extracting, transforming and mapping data from one format into another format and / or location. The objective is to make the data of high quality and suitable for downstream purposes like analytics. Very similar to the ETL (extract, transform and load) and ingest concepts.

EWSolutions™

- ❑ **Signal:** is the meaningful information in the data set
- ❑ **Noise:** is the random, unwanted variation or fluctuation that distracts from the relevant information (Signal)
- ❑ **Data Democratization:** is a principle that states that data should be available to everyone in each organization, not just key analysts, decision-makers, or leaders. It is tied to concepts like self-service and service architectures. High levels of data literacy and data governance are needed for this concept to be successfully implemented.
- ❑ **Frequency:** is the number of occurrences of a repeating event per unit of time. For example, how many tables purchased were blue for each month of a year.
- ❑ **Correlations:** A relationship or connection between two things (signals) based on co-occurrence or pattern of change.

Explanatory Analysis vs. Exploratory Analysis

❑ **Exploratory Analysis:** is what a data scientist or data analyst does to get familiar with the data

- ➤ A data scientist may start out with a hypothesis or question or may be mining the data to determine what might be interesting to explore further
- ➤ Some use the analogy of turning over 100 rocks to find 1 or 2 precious gemstones

❑ **Explanatory Analysis:** is what happens when a data analyst or data scientist has something specific to want to present to the business users

- ➤ Storytelling with data is about explanatory analysis and how most effectively to do it
- ➤ In our previously analogy this would be presenting the 1 or 2 precious gemstones to the business users

Standard Deviation

❑ **Standard Derivation:** is the average amount of variability (difference) in your dataset. On average, how far each value lies from the mean

❑ Standard derivations can be less than 1. For example, suppose you are analyzing the number of apples eaten by kids each day by county in Illinois. The standard derivation between the counties is most likely below 1

❑ **High Standard Deviation:** occurs when the values are far from the mean

❑ **Low Standard Deviation:** occurs when the values are clustered close to the mean

Normal Distribution

☐ **Normal Distribution (or Normal Curve):** occurs when data is symmetrically distributed with no skew

➢ Common examples include height, standardized test scores, etc..

➢ Has a symmetrical "bell" shape. Fold the curve in half and the two sides match perfectly

☐ Mean is equal to the median

☐ In a normal distribution, 68% of the data falls within -1/+1 standard deviation from the mean, and 95% of the data falls within -2 and +2 standard deviations from the mean
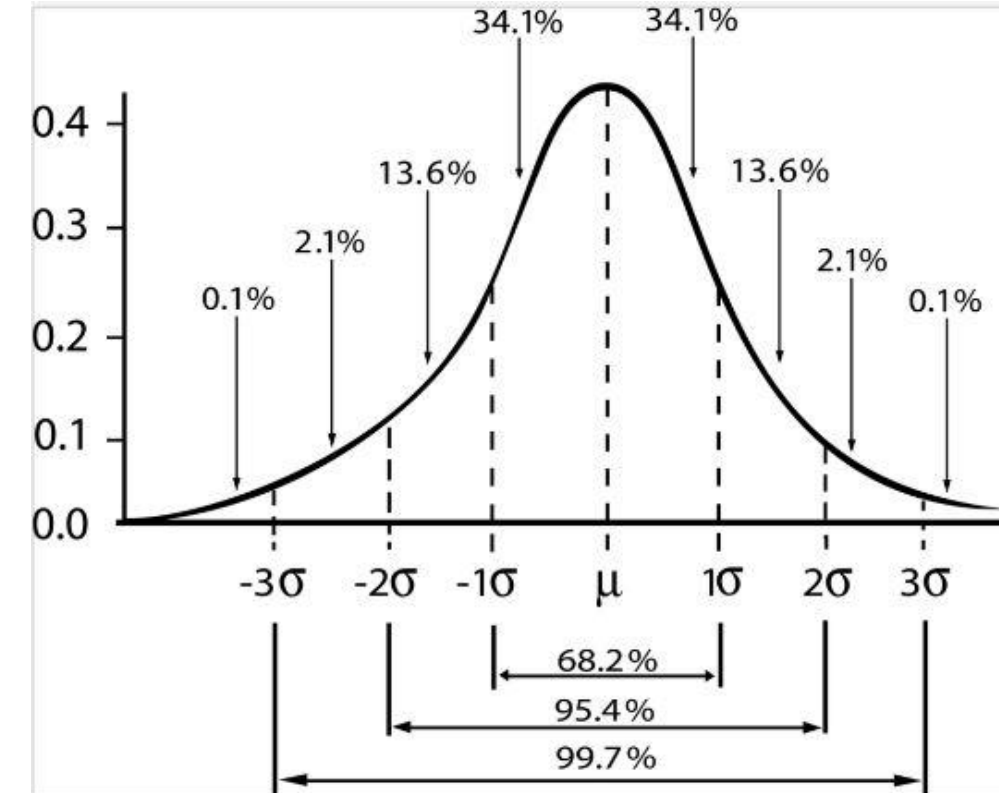
EWSolutions™

Data Management Consulting Services

Over 150 Client Partners
100% Data Management Success since 1997

GOVERNANCE METADATA ANALYTICS
EWSolutions METHODOLOGIES
EWSOLUTIONS.COM

World-Class Training

More than 40 courses, taught by industry experts. Over 160 satisfied client partners

EWSolutions™

DATAMANAGEMENT EDUCATION.COM

Online, On-Demand Courses

Data Warehouse Data Models

DataManagementU.com
Free knowledge portal

Big Data Meta Model