



Data Quality
*When is Good Enough, **Enough?***

Data Quality
*When is Good Enough, **Enough?***

***When it is good
enough!***

- Increase your understanding that there are no absolutes when it comes to **Data Quality**
- To understand that:
 - there only shades of **Gray**
 - That the level of Quality is dependent upon what we define **RIGHT** to be
 - Higher levels of Quality cost more **\$\$\$**

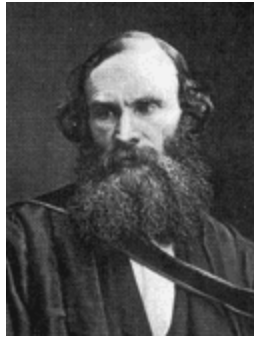
**We may never know or be able to achieve
100% correctness**

Goals & Objectives Of Presentation

- Increase your understanding that there is no absolutes when it comes to Data Quality
- To understand that:
 - there only shades of **Gray**
 - That the level of Quality is dependent upon what we define RIGHT to be
 - Higher levels of Quality cost more **\$\$\$**
- We may never know or be able to achieve 100% correctness – ***live with it I do!***

To measure is to know.

"If you can not measure it, you can not improve it."



Lord Kelvin

***Quality is everyone's
responsibility***

W. Edwards Deming

- *Quality is everyone's responsibility*
W. Edwards Deming
- ***Quality is the result of a carefully constructed cultural environment. It has to be the fabric of the organization, not part of the fabric***

Philip Crosby

- To know it, we have to:
 - be able to Measure it,
 - Understand it, and
 - It is my responsibility
 - And its cultural

Still does not tell me what it is!

Why is Data Quality like porn?

- **You know it when you see it!**

paraphrase of, Justice Potter Stewart of the United States Supreme Court 1964

Or is there an actual definition and set of metrics?

A Working definition of Data Quality:

- *The data is accurate and precise enough to do what I need to do!*
- *What I need to do, knowing what has to be done, is the Problem*
- *And what needs to be done varies by person and function*

Phil Teplitzky 2010

and there is the rub,

- Knowing what people want and mean,
- But there are a lot of people in an organization,
- That means figuring out what is different and what is the same
- *We cannot understand each other when we talk how do we add precision and accuracy to Market and Meta Data?*

- Closing price? Opening Price?
 - Asked, bid, settled?
- Date and Time?
 - Local time, passing the book, multi-part transactions
- Counter Party?
 - Name, legal name, owner, subsidiary of,
- Credit Facility exposure – need to know organizational structure
- Corporate Actions?
- Rounding errors

- What is an inch, a meter, a pound a kilogram?
- Is a kilogram the same unit of measure as a pound?
- What is round?
- In architectural drawings they always give the tolerance – 3 feet 6 inches plus or minus .01inches

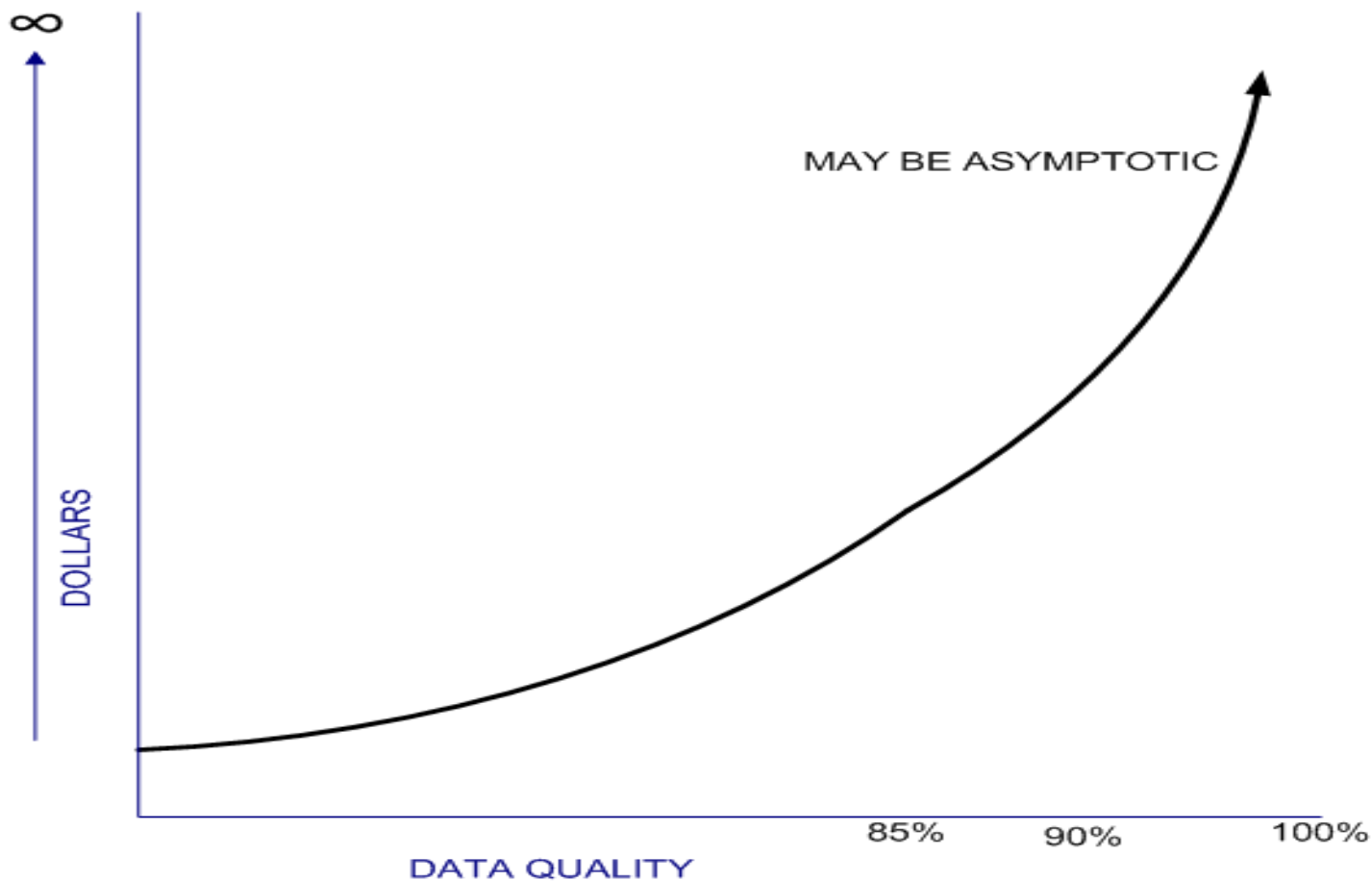
Dimensions of Quality

Dimension	Definition
Accessibility	The extent to which data is available or easily and quickly retrievable
Appropriate Amount of Data	The extent to which the volume of data is appropriate to the task at hand
Believability	The extent to which the data is regarded as true and credible
Completeness	The extent to which data is not missing and is of sufficient breadth and depth for the task at hand
Concise	The extent to which data is compactly represented
Consistent Representation	The extent to which data is presented in the same format
Ease of Manipulation	The extent to which data is easy to manipulate and apply to different tasks
Free of errors	The extent to which data is correct and reliable
Interpretability	The extent to which data is in appropriate languages symbols and units and the definitions are clear
Objectivity	The extent to which data is unbiased and unprejudiced and impartial
Relevancy	The extent to which the data is applicable and helpful for the task at hand
Reputation	The extent to which data is highly regarded and in terms of its source or content
Security	The extent to which access to data is restricted appropriately to maintain its security
Timeliness	The extent to which the data is sufficiently up to date for the task at hand
Understandability	The extent the data is easily comprehended
Value Added	The extent to which data is beneficial and provides advantages from its use

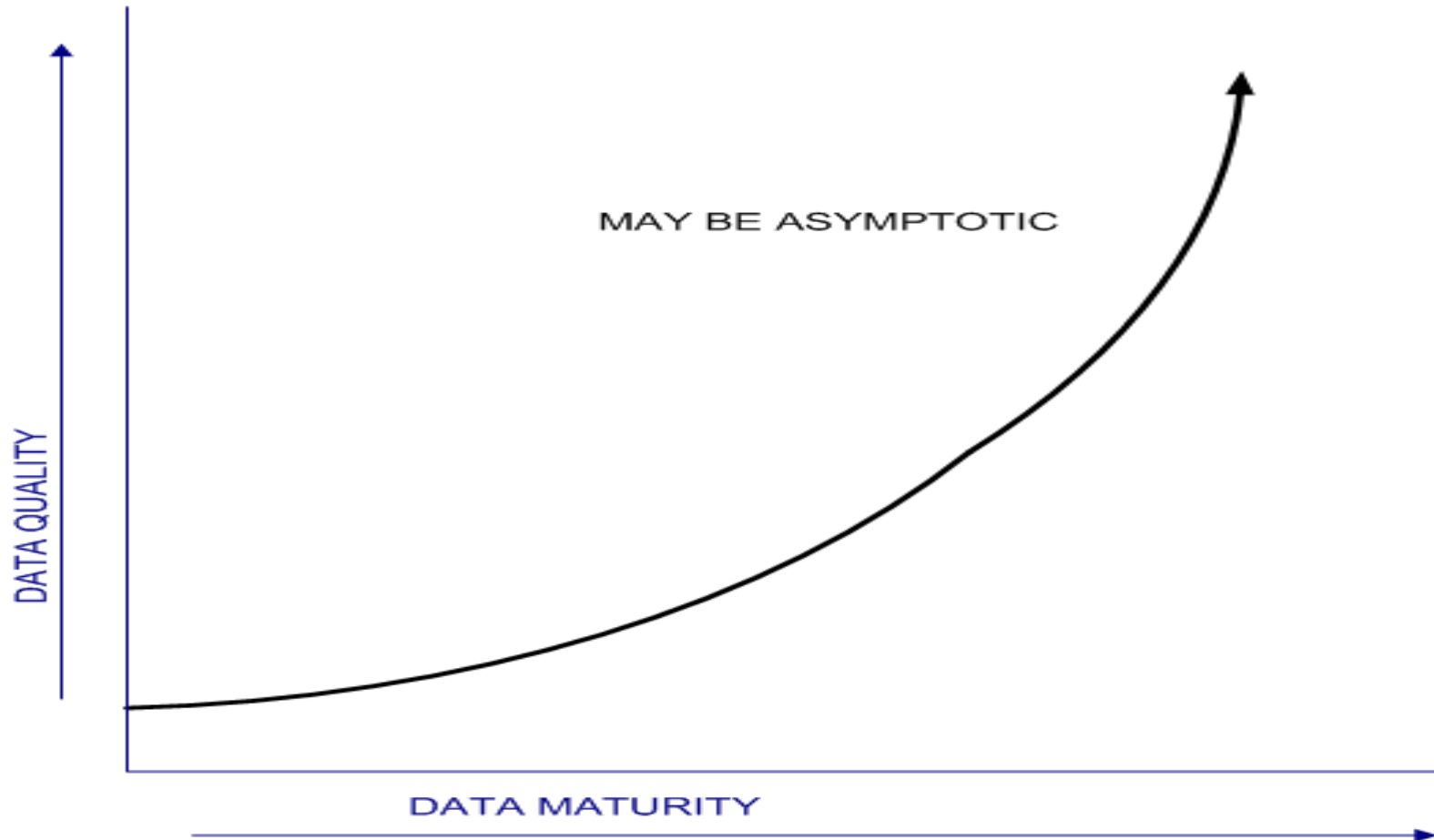
So what does this all mean?

1. That data is relative
2. That achieving 100% quality may be impossible or else **really** expensive
3. That it cannot be achieved by itself but only within the context of an overall data maturity framework

Financial Implications



Data Maturity vs. Data Quality



- If Quality is a function of our definition of Correct
- And if our definitions are based on language, which by its very nature is imprecise (see the multiple definitions of common English words) then how accurate, complete and precise can the definitions of the characteristics and attributes used to create the correct definitions?
- You can only be as accurate and precise as the definitions you provide

- Data Quality observes the rules of Quantum Mechanics more than the rules of Newtonian Mechanics!
- It depends on the observer, and it can only be **statistically measured!!!!**
- By and large the only thing we can say for certain is- that if there is an error in Data Quality then there are more!
- If you have not found any – you just have not looked hard enough!

- Closing price – Bid, asked, settled?
- Interest calculations – do you really need them to the 5 decimal place?
- Greeks and Curves – how many data points what algorithm should you use?
- Counter Party and Customer Hierarchies – 50% rule, any ownership, recursive ownerships?

Data Quality like Zeno's Paradox

- You can never get to 100% but you are close enough for all practical purposes!

- So be a good Quality Engineer and define for good ENOUGH !

**AT LEAST FOR WHAT YOUR PURPOSE
IT**