



Data Quality

IOSCO Conference 2016 Workshop

May 10, 2016

Srinivas Bangarbale

Chief Data Officer

U. S. Commodity Futures Trading Commission





Disclaimer

The views expressed in this presentation and discussion are the personal views of Srinivas Bangarbale and do not necessarily reflect the views and or positions of the Commodity Futures Trading Commission, its Chairman, Commissioners or staff.



Agenda

The problem

- What we are trying to solve

Status of current work

- What current initiatives are underway

Data Quality

- How data quality is defined

Data Quality in the context of a financial regulator

- Why is data quality key for financial regulation

Thoughts on Data Quality improvements

- What can be done to get closer to solving the issues



THE PROBLEM



Regulatory issue

The problem during the 2008-09 crisis from a regulator's perspective

- Lack of comprehensive data on Over the Counter Derivatives (OTCD) transactions and positions
- Lack of high quality data that provided a clear view of risks
- Lack of standardized data to aggregate globally and understand systemic impacts
- Lack of timely availability of data to mitigate market abuse and risks



The data problem today

- Regulators have been collecting OTCD data for over 3 years
 - But data continues to have consistency and quality issues
 - Data reported using different standards on different legs of cross-border trades
- Regulators have made progress in removing barriers to share data, but further work needs to be done
- Global OTCD data need to be aggregated by regulators individually
 - Substantial policy and technical work need completion to reach the goal
 - Other regulatory proposals such as global aggregation of data



How Swaps differ from Futures

- More complex contract terms
- Significant portion of the market is *bespoke* – custom contracts
- Long lifespan - many events reported over lifetime (*Continuation Data*)
- Global reach of contracts – many Swaps are inter-jurisdictional (*cross-border*) in nature
- Ongoing and higher-degree of innovation in Swaps products



STATUS OF CURRENT WORK



G20 and regulatory commitment

- All OTCD contracts must be reported to Trade Repositories (TRs)
 - Provide a comprehensive view of the markets and market activity globally
 - Understand systemic risk by global aggregation of data reported to TRs
 - Harmonize data so that global aggregation is possible (e.g., UPI, UTI, LEI)

While this is a first and critical step to understand the functioning of the global OTCD markets, this is not enough



Aggregation and Harmonization

- Legal Entity Identifier (LEI) is a success story
- Financial Stability Board (FSB) Aggregation Feasibility Study Group (AFSG)
- Committee on Payments and Market Infrastructures (CPMI) and International Organization of Securities Commissions (IOSCO)
 - Currently developing guidance on Unique Transaction Identifier (UTI), Unique Product Identifier (UPI), and other data elements needed for global aggregation
- OTC Derivatives Regulators Forum (ODRF) Data Harmonization



Lessons learned from Futures and Options data

- CFTC has been receiving Futures and Options data since 1975
- “Rome wasn’t built in a day”
 - Data will not be perfect on day one, but evolves over time through iterations, vigilance and effort
 - Each format change (e.g., 80 byte records to FIXML) takes significant time and effort to iron out
 - Time deepens understanding – data validations have gotten more sophisticated over time
- Standards - “Without standards, there can be no improvement” – Taiichi Ohno
 - Best to use consensus, industry-wide / international standards that firms already use
 - Engage industry in standards setting from the beginning
 - Mayhem without standards – When propriety standards were set by exchanges for trade register data, CFTC spent significant effort harmonizing the data in-house



Lessons learned from Futures and Options data (continued)

- “An ounce of prevention is worth a pound of cure”
 - The more time you spend on developing a data spec, the less time you spend cleaning up data
- “The only thing that is constant is change” - Heraclitus
 - Markets continually change, evolve, and innovate
 - Design today but keep flexibility for tomorrow
- “You get what you pay for”
 - Data quality takes resources
 - Cannot achieve data quality by sub-optimal investment



DEFINITION OF DATA QUALITY



What is *Data Quality*?

- Data Quality is the availability of the *right data*, at the *right time*, in the *right format*
- The most wonderful data not fit for purpose is useless
 - E.g., collecting the street address of the legal entity with **every transaction** will not help in calculating the net exposure of the counterparties
- Every piece of data must meet the requirements of relevance, standard, and timeliness

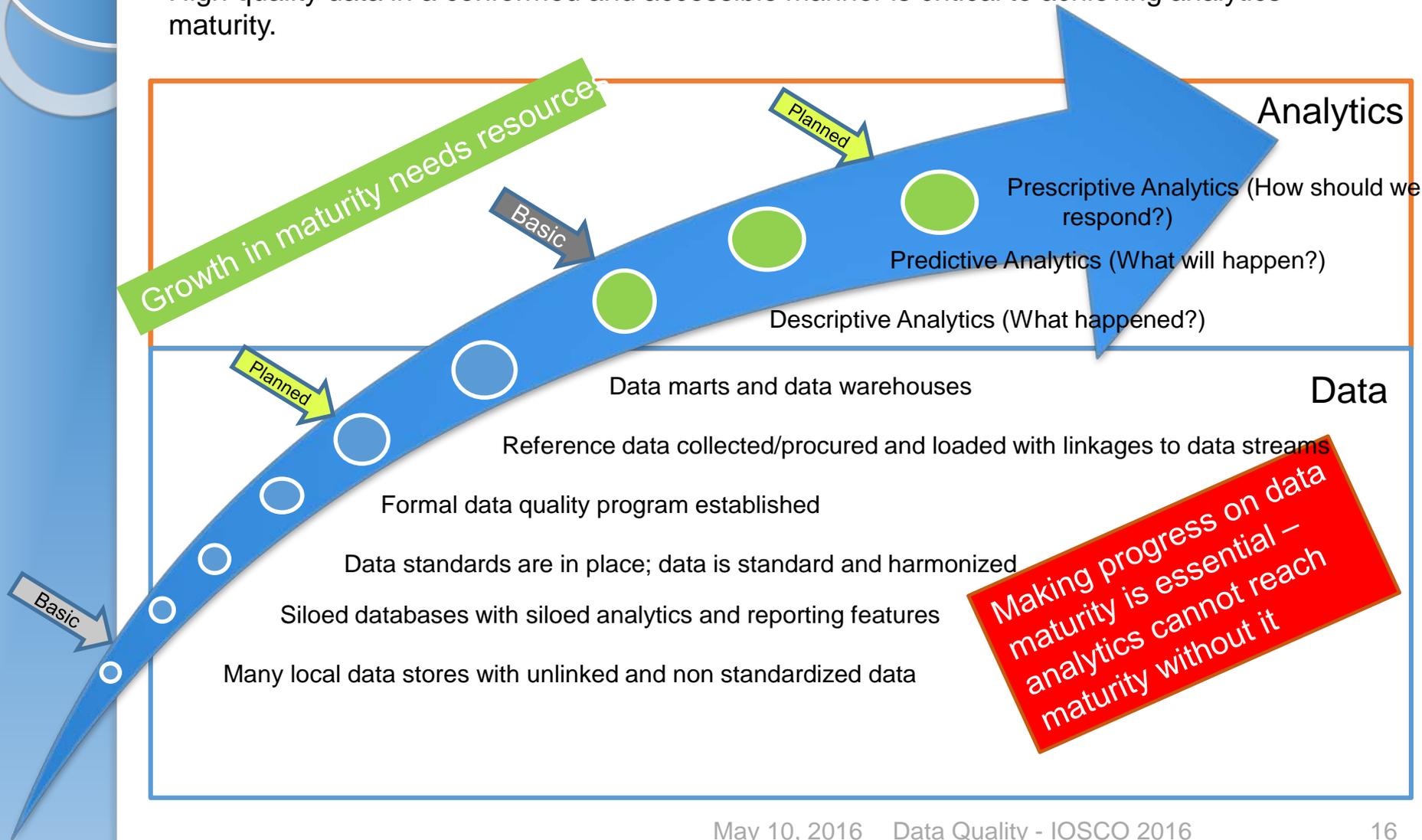


Components of Data Quality

- **Relevance** – only necessary data
 - Data not necessary for use cases identified for individual regulator's functions or global aggregation - creates noise in the system and wastes resources to maintain needless data
- **Standard** – standardized data for all data reporting
 - Data not standardized (definition, form and manner) – hinders effective aggregation across trade repositories and globally
- **Timeliness** – neither too early nor too late
 - Data not available at the right time will not allow us to obtain insights into market fluctuations - hinders predictive capabilities necessary to mitigate risks in a timely manner

Maturity path for data use

High quality data in a conformed and accessible manner is critical to achieving analytics maturity.



Detriments to data quality

- Unclear rules
- Rules created without clear use case definition, agreement, and documentation

Rules

- Too many standards
- Standards that are not fit for purpose
- Standards that are not mandated

Standards

- Disparity in technology between regulators and market participants
- Insufficient technical platforms for smaller market participants

Technology

Skills

- Lack of financial data experts in the beginning of the process
- Lopsided expertise balance between legal data standards



RECOMMENDATIONS FOR IMPROVEMENT



Recommendations

- Tackle data quality methodically
- Begin with the end goal in mind
- Issue clear guidance on
 - Relevance - what data is needed
 - Timeliness - when it is needed
 - Standard - how it should be submitted
- Plan for data anomalies
 - Consider capabilities necessary to respond to anomalous data
- Develop SLAs for TRs and data sharing partners
- Strictly enforce compliance in data reporting



Tackle data quality methodically

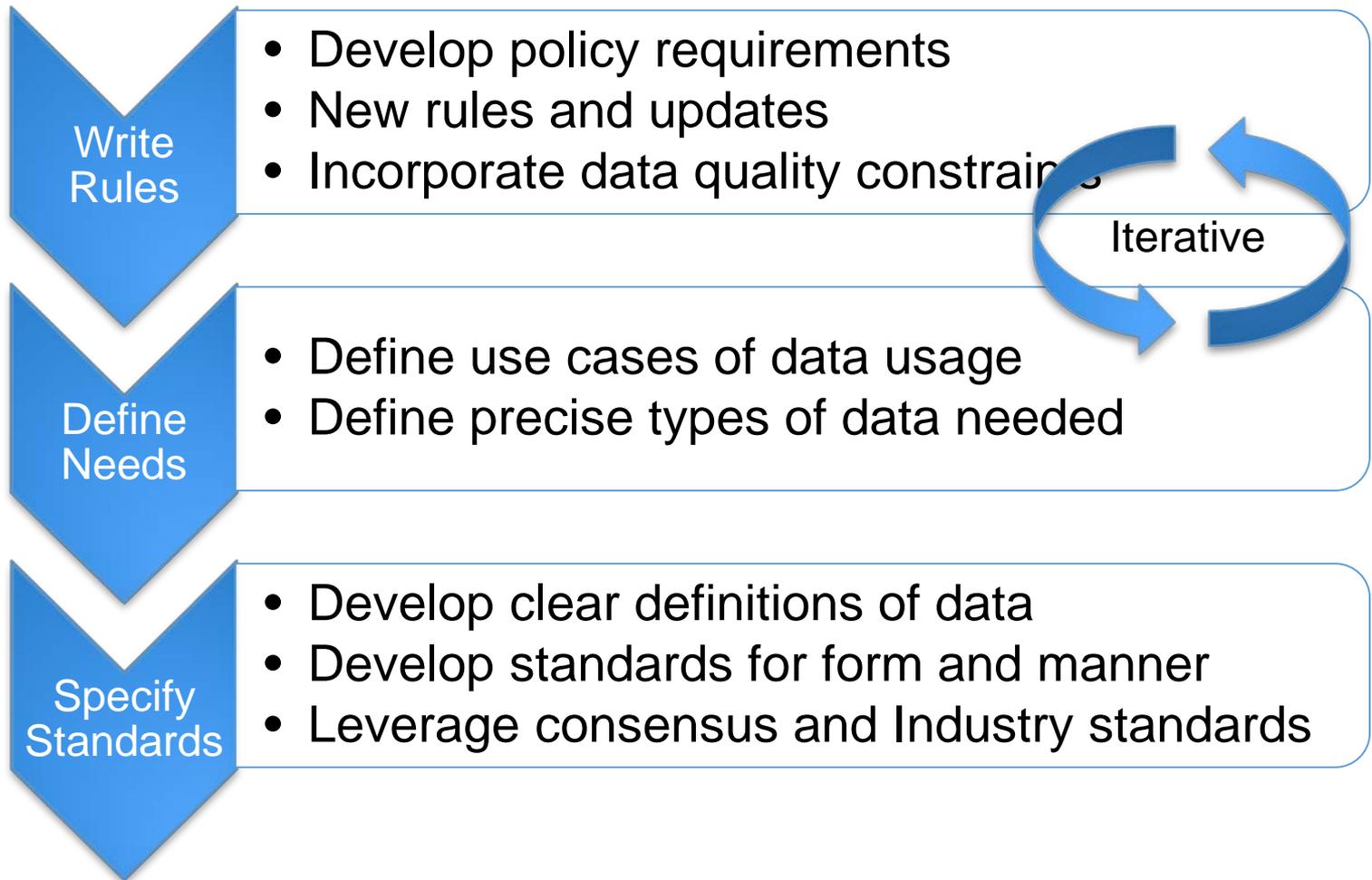
Definition Define data needs methodically

Acquisition Develop the structure to obtain and house the data

Use Develop the ability to use the data the right way for the right purpose



Data stages – *Define* (1 of 3)





Data stages – *Acquire* (2 of 3)

Architect Data

- Develop proper data models
- Understand the interrelationships among data sets

Ingest Data

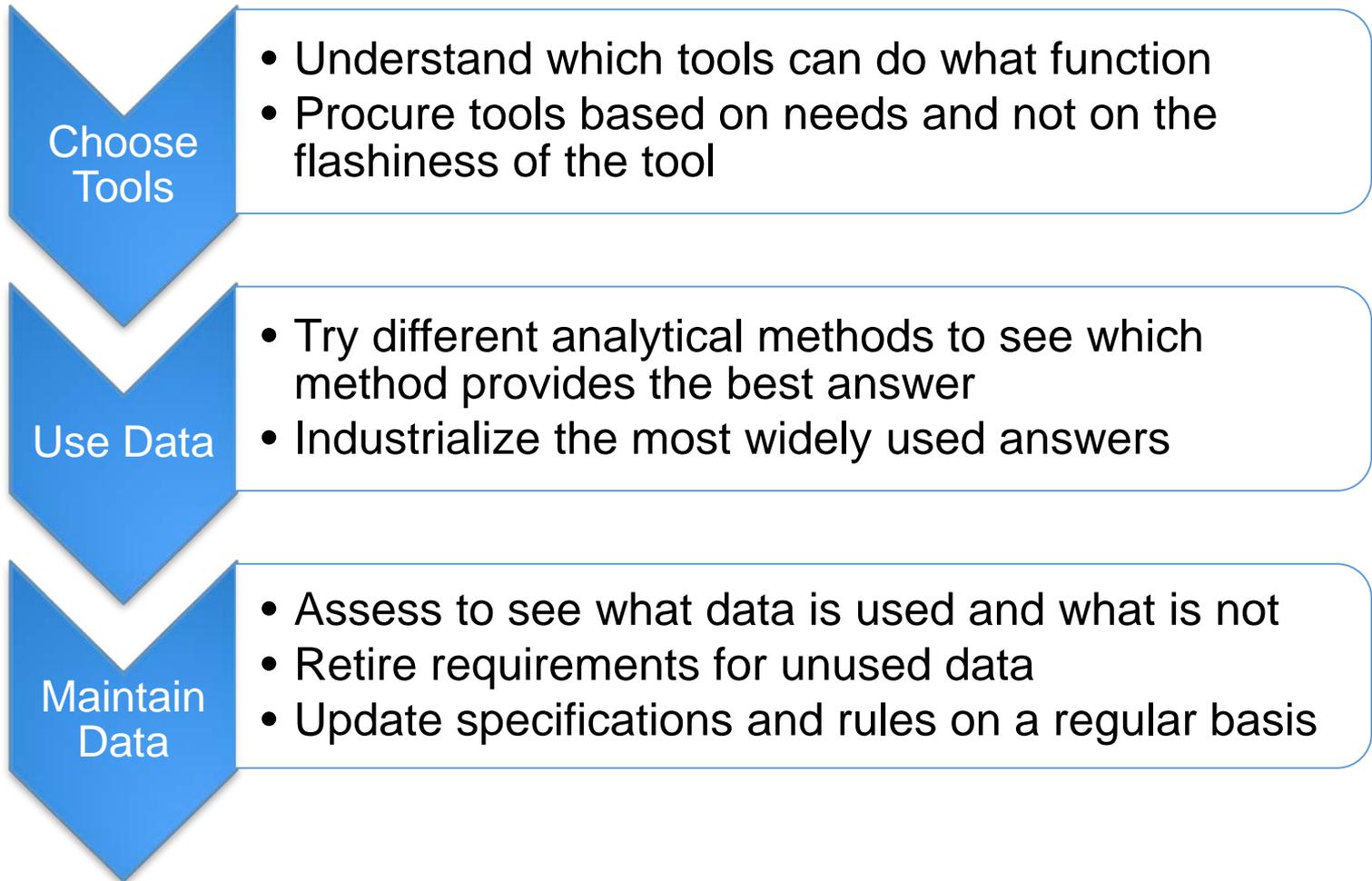
- Validate data as it is loaded into data stores
- Provide users the ability to verify initial data ingest

Enhance Datasets

- Acquire additional reference and value added data
- Link core data sets with value added data sets to provide enriched information



Data stages – *Use* (3 of 3)





Begin with the end goal in mind

- Identify the right set of precise use cases
 - Loose definitions like “understanding global risk” are insufficient
 - Deeper dive on the use cases along with a common understanding of the problem the use case is solving
 - For example, include a basic understanding of how risk is calculated and the various dimensions of risk
 - Mapping rules across regimes as opposed to merely mapping data elements across regimes helps understand gaps in data across jurisdictions
- Use dictates data need



Provide clear guidance

- Provide a clear and unambiguous list of what data you need
 - Beyond providing mere rule text for data provide clear guidance on what data your need and clear and consistent definitions
- Provide clear timeliness requirements for reporting data
- Identify clearly the standards to use for reporting data
 - Beyond form and manner, provide clear and consistent definitions of data - form and manner cannot compensate for fuzzy definitions
 - Merely asking for “Execution Timestamp” and specifying ISO 8601 as the standard will be unhelpful if the regulators don’t define what “Execution” means



Plan for data anomalies

- Identify response mechanisms and thresholds
 - *What* should be done when thresholds are reached?
 - *Who* will be alerted and what information do they need to make decisions?
 - *What* type of enforcement actions can be taken?
 - *What* other data are needed to enable this decision making?
- Sometimes the most stringent use cases satisfy 80% of basic requirements



Develop SLAs

- Develop appropriate service level agreements (SLAs) for TRs and between parties sharing data
 - How quickly is each piece of data needed?
 - What feedback is necessary to provide this piece of data? How often should the feedback be given?
 - How can it be assured that each party meets their SLA (Operational Governance)?



Enforce compliance in reporting

- There is virtue in compliance
 - Essential to performing the regulatory mission
- Reporting accurately and on time is not a favor
 - Parties and market participants have a clear and affirmative legal obligation to report accurately and on-time
- Violations of such obligations need to be firmly dealt with
- Excusing poor behavior by reporting parties tends to encourage poor quality data reporting



Maintenance and agility

- Maintenance of data specifications
 - Regulatory data specifications have to change with market practice innovations
 - Maintenance is essential to continued data quality
- Agility in reaction
 - Critical to ensuring continued high quality
 - Keeping up with shifts in market practices, technologies and standards is not a part-time job



° **BACKUP SLIDES**

Futures & Swaps
Position Data
16M / Day

Metrics

Data Feeds

(approximate # of million **records** per day
CFTC is currently loading)

Futures Intraday
Trades /
Swaps Events

10M / Day (Note: adding order
messages would be orders of
magnitude greater, adding 200
– 300M records per day.
At scale, adds another
7 - 10 feet to
the diameter)



FERC
Data
0.33M / Day



Market Data
0.2M / Day

Scale: 2.5M records / inch



Margin
& Collat

Margin &
Collateral
3M / Day



SDR Data
(not to scale)



Valuation / Stress Test
0.1M / Day



Fin Statement / Capitalization
0.1M / Day



Org / Product Data
0.1M / Day

Time & Sales Futures
Data

(top of the order book)

240M / Day

(not to scale, at full scale
this bubble would be 8 feet across)

Enforcement
.2 – 1TB / Week
(not to scale, scale = 10 feet)