

Why business is not interested in data quality

April 2009



Agenda

- Background to Shell and the oil industry
- Perceptions of data value
- Quantify the value of data
- Quantify the cost of data quality





Our business – Upstream

- Explore
 - Geophysics, geology, wells
- Develop
 - Reservoirs, joint ventures
- Produce
 - Platforms, crudes
- Maintain Assets
 - Materials, equipment



Our business – Downstream

- Supply and Trading
 - Crudes and intermediates
- Manufacture
 - Refinery, lube oil blending, packed products
- Distribute
 - Inland transport
- Sell
 - Companies e.g. mining, road transport, airlines, shipping, military...
 - Retail, Distributors, Card holders
- Maintain Assets



The nature of the oil industry

- Oil is a commodity
- Downstream margins are low
- Demand is growing beyond the ability to supply cheaply
- Products need to be handled safely



Our business imperatives

- More upstream, profitable downstream
 - Find and produce more crude oil and gas
 - Manage the portfolio and control costs
- Safety
 - Do no harm to people
- We aim to be consistently a top quartile performer
- No mention of data or data quality



We think data quality can help

- Assess prospects
- Reduce costs
- Define risks to safety and security and support controls

But can we prove it?



We have heard it all before

- Single source of the truth
 - At what cost?
- Enterprise data model
 - Who does it really help?
- Data -> Information -> Knowledge -> Wisdom
 - We' ll settle for the data right now
- Facts based decision making
 - We have no facts on data quality
- Leveraging data for business growth
 - But we want to contain costs
- Data is an asset
 - What is it really worth?
- We live in an information age
 - Its still GIGO

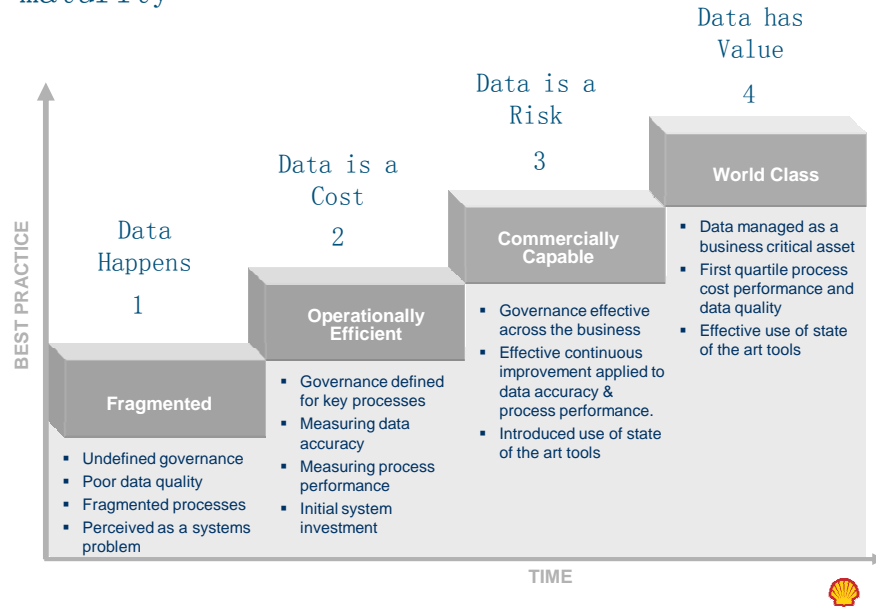


Agenda

- Background to Shell and the Oil industry
- Perception of data value
- Quantify the value of data
- Quantify the cost of data quality



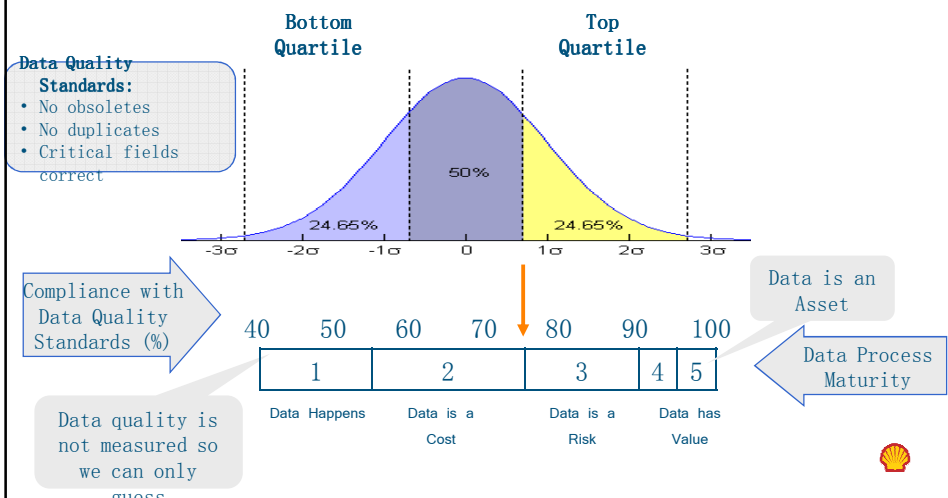
Perceived value of data changes with maturity



Top-Quartile Data for the business

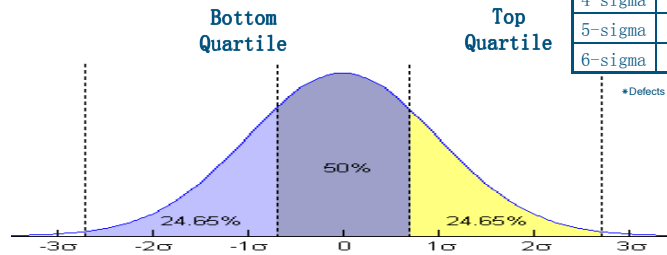
Gartner group Strategic Planning Assumptions:

“Through 2010, more than 75% of organizations will not get beyond Levels 1 and 2 in their data quality maturity (0.8 probability)”.



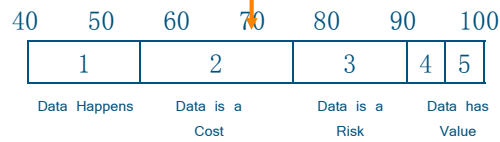
What is Top-Quartile Data today?

Level	DPMO*	Yield
1-sigma	690,000	31.00%
2-sigma	308,000	69.20%
3-sigma	66,800	93.32%
4-sigma	6,210	99.38%
5-sigma	233	99.98%
6-sigma	3.4	100.00%



*Defects per million opportunities

Compliance with Data Quality Standards (%)



Data Process Maturity



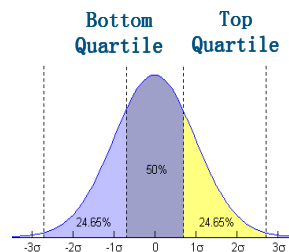
Outlook for 2012

Gartner group Strategic Planning Assumptions:

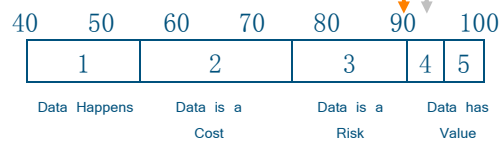
“Through 2012, less than 10% of organizations will achieve level 5 data quality maturity (0.8 probability).”

Level	DPMO*	Yield
1-sigma	690,000	31.00%
2-sigma	308,000	69.20%
3-sigma	66,800	93.32%
4-sigma	6,210	99.38%
5-sigma	233	99.98%
6-sigma	3.4	100.00%

*Defects per million opportunities



Compliance with Data Quality Standards (%)



Data Process Maturity



DQ professionals are an endangered species

- Data is viewed as an unnecessary cost to most companies when times are tough
- Companies think that IT has solved their data problems
- The promise of data quality is rarely fulfilled because data is managed as 1's and 0's - not as an asset
- DQ professionals are scarce
- No internationally recognised qualification



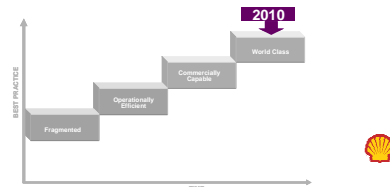
Agenda

- Background to Shell and the Oil industry
- Perceptions of data value
- Quantify the value of data
- Quantify the cost of data quality



Quantify the Value of Data

- Valued by “Scrap and Rework”
- Demand for tangible value of data quality
 - Understand impact of data quality on business processes
- Focus on items under the radar screen
 - Not detected by controls
 - Business scrap and rework
- Value comes from processes outside data management
- Measurement of quality provides the proof

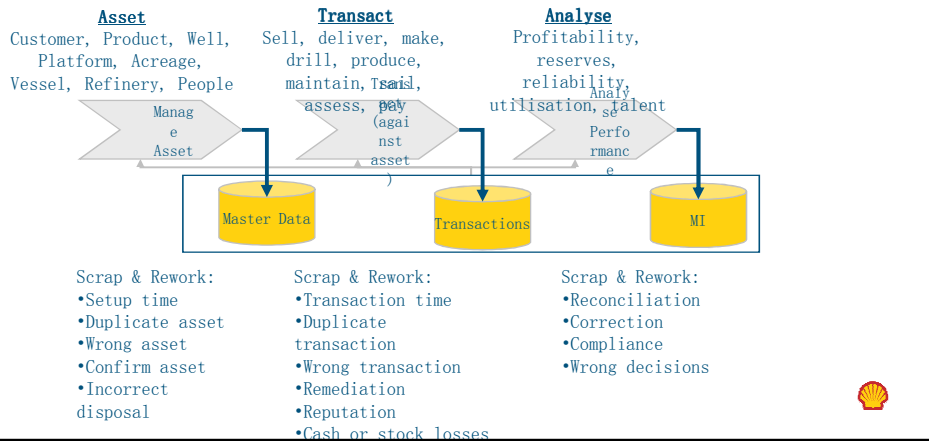


Focus on master data

The only data unique to a transaction are: Date, Quantity and Unit Rate. Everything else is a reference to master data.

Date	Quantity	Unit rate	Customer	Product	Profit Centre
------	----------	-----------	----------	---------	---------------

Each master data record, and each defect on a master data record, impacts ~ 10 to 100 transactions per year.



Valuing data by Scrap and Rework

- Complements continuous improvement
- Consider the impact of poor quality data on each transaction
 - Each data error is multiplied by volume of transactions
- Assume all transaction errors are detected and corrected i.e. scrapped and reworked
 - Tangible “best case scenario” cost rather than potential cost avoidance
 - Includes errors that slip through “train wreck” based controls



The Value of Data Equation

Cost of Scrap and Rework

=

V_t

times

C_r

times

$(1 - Q)$



Value of Data Equation

- V_t = Annual volume of relevant transactions per object
 - e.g. 1 million orders for 10,000 vendors = 100 orders per annum per vendor
- C_r = Cost of rework per transaction in error
 - e.g. Typically \$50 to journal and re-post a transaction
- $(1 - Q)$ = Data quality defect rate
 - From measurement against relevant data quality standards

**You must measure data quality before
you can estimate the value of data in
preventing scrap and rework**




Case Study – Vendor Classification


- Many of our materials are on contract. Knowing the material you need takes you to a contract and hence the vendor
- Materials off contract require selection of a vendor who supplies the class of material. Knowing the class of material allows you to search for a vendor who supplies that class.
- Vendor selection takes less than 5 minutes if the classification is present and correct.



Case Study – Vendor Classification

- If the classification is missing, vendor selection takes 3 or 4 hours trawling through data and cross-referencing
 - % of unclassified vendors x volume of purchases off-contract x cost to find vendor = \$\$\$
- If the classification is present but incorrect you end up raising an order against the wrong vendor leading to scrap and rework plus another search for the right vendor. Lost time, rework effort, new search effort.
 - % of incorrectly classified vendors x volume of purchases off-contract x (cost to scrap order + cost to find vendor) = more \$\$\$
- At what point does the user ignore the classification because it cannot be trusted? 
 - Assuming 1 in 3 means that if the quality of vendors

Obstacles to Progress

- Users of data compensate with their own knowledge
 - *“The user knows which data to use”*
 - Job for life but how to manage succession?
- Controls prevent the worst problems/ train wrecks
 - Death of a thousand cuts
- We reward fire-fighters – people who fix problems rather than prevent them e.g. Incorrect master data causes 50 transaction errors per year
 - Person who detects and fixes each error quickly gets praised 50 times per year
 - Person who fixes the master data regularly so no errors occur has no opportunities for praise
- Sledge hammers/ Brute force
 - If its important enough we make it work e.g. closing 

Obstacles to Progress

- Automation and off-shoring reduce the cost of rework
 - Changes the break-even point but there is still a break-even point
- Data quality not measured or data quality measures are disputed
 - Data assumed to be OK - *“Our processes are working”*
- Poor data quality is accepted as reasonable
 - How good is enough?



Agenda

- Background to Shell and the Oil industry
- Perceptions of data value
- Quantify the value of data
- Quantify the cost of data quality

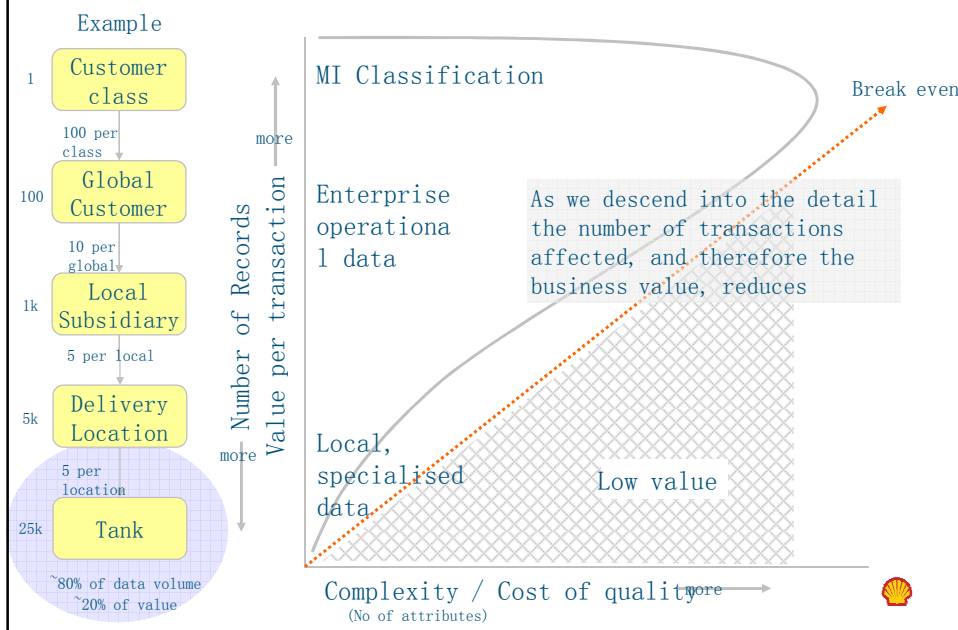


Cost to ensure data quality

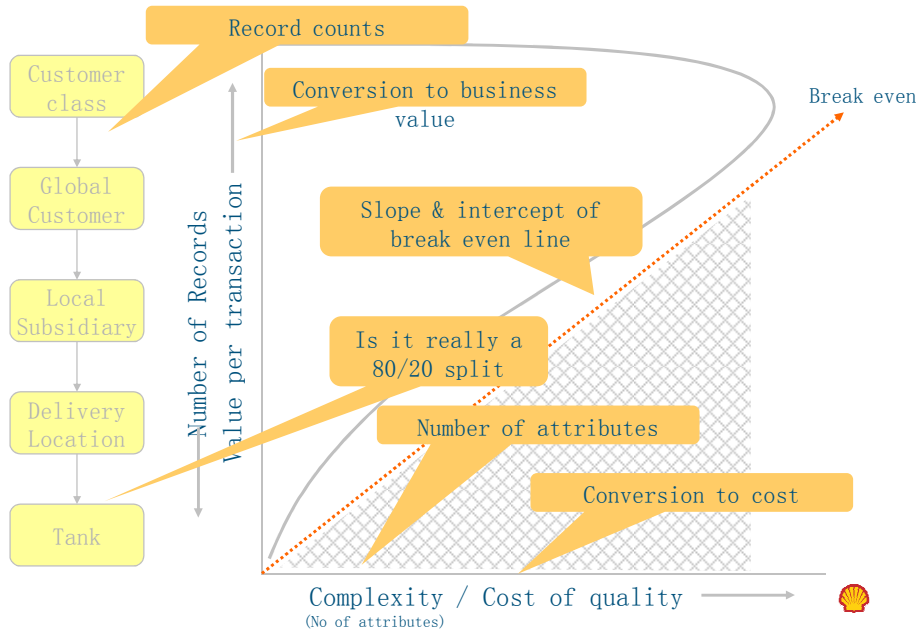
- Validate and maintain with full time, qualified professionals
 - Vs average 20% of time spent on data
- Manage the data process end-to-end
 - Gate-keeping to ensure approvals and performance
- Measure data quality and process performance
 - Perform QA, collect metrics
- Continuous improvement of data processes
 - Collect issues, root cause analysis, improvement projects



How much to spend on quality?



What we need to know



What we still need

- What should we be measuring to quantify data quality?
- How do we measure it efficiently and effectively?
- What is the top quartile process maturity benchmark?
- What is the top quartile data quality benchmark?
- What is top quartile data management process performance?
- Learning programmes and certification



Conclusion

- Business focus is not on data quality per se
- Business perceives data as a cost
- Quantifying the value of data quality is critical to making progress
- Quantifying the cost of data quality is critical for sustainability
- We are short on facts
- There is still a long way to go

