Data Quality Metrics

Status of the National Data Repository (NDR) work for Regulators - 2014-2017

Philip Lesslar

Digital Energy Journal Conference
3rd October 2017
Kuala Lumpur

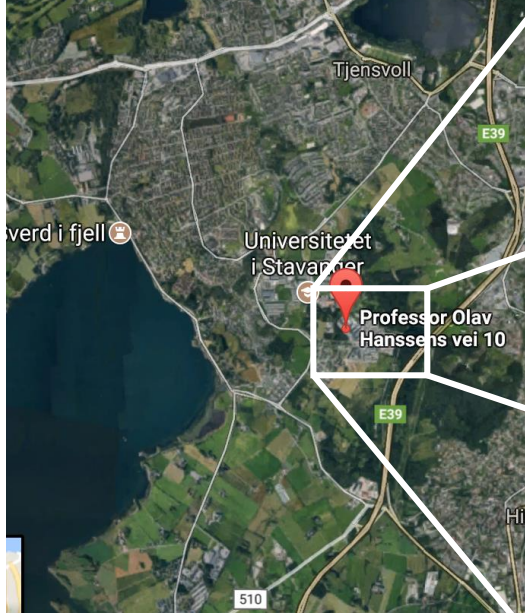# Summary of NDR 2017, 6-8 June, Stavanger, Norway
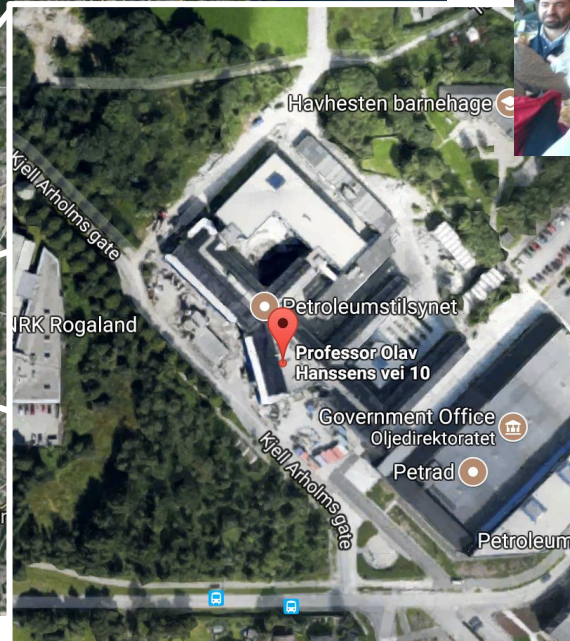
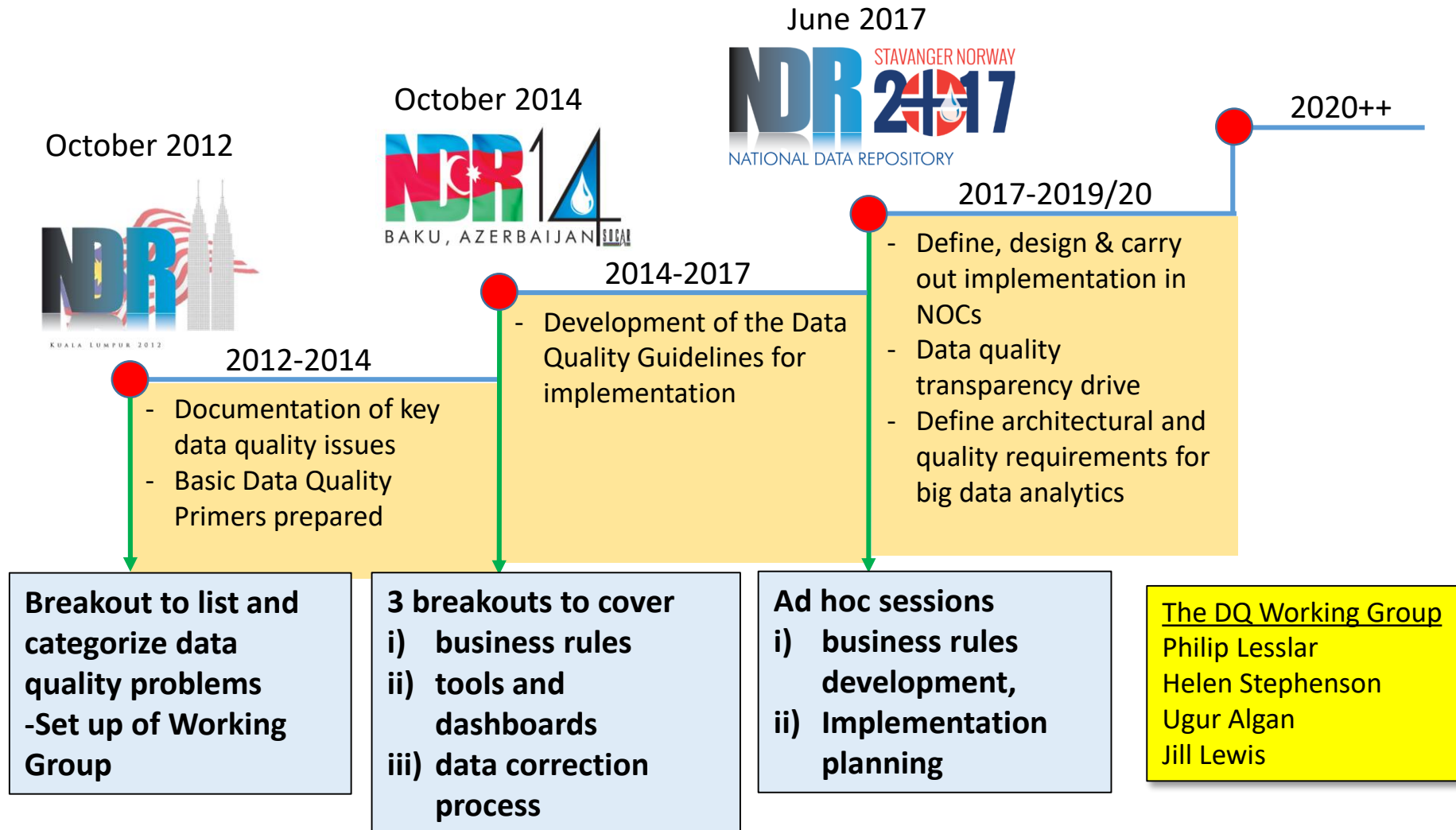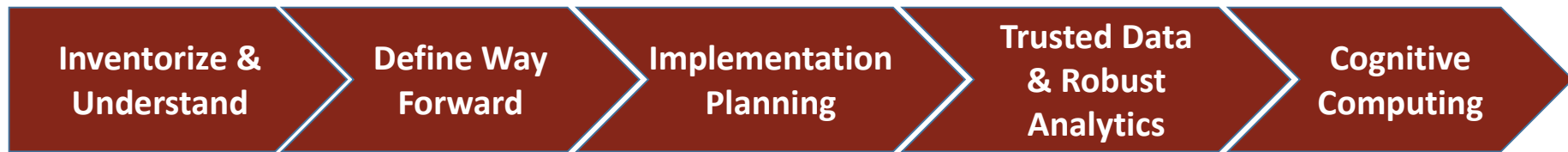Sponsored by:

NORWEGIAN PETROLEUM DIRECTORATE

Organized by:

energistics® Energy Standards

- 165 participants
- Representing 30 countries
- 20 sponsors

Google Maps

Data Quality Metrics – The Journey

Inventorize & Understand ▷ **Define Way Forward** ▷ **Implementation Planning** ▷ **Trusted Data & Robust Analytics** ▷ **Cognitive Computing**

October 2012

October 2014

June 2017

2020++

2012-2014
- Documentation of key data quality issues
- Basic Data Quality Primers prepared

2014-2017
- Development of the Data Quality Guidelines for implementation

2017-2019/20
- Define, design & carry out implementation in NOCs
- Data quality transparency drive
- Define architectural and quality requirements for big data analytics

Breakout to list and categorize data quality problems
-Set up of Working Group

3 breakouts to cover
i)   business rules
ii)   tools and dashboards
iii) data correction process

Ad hoc sessions
i)   business rules development,
ii)   Implementation planning

The DQ Working Group
Philip Lesslar
Helen Stephenson
Ugur Algan
Jill Lewis

Data Quality in the context of NDRs
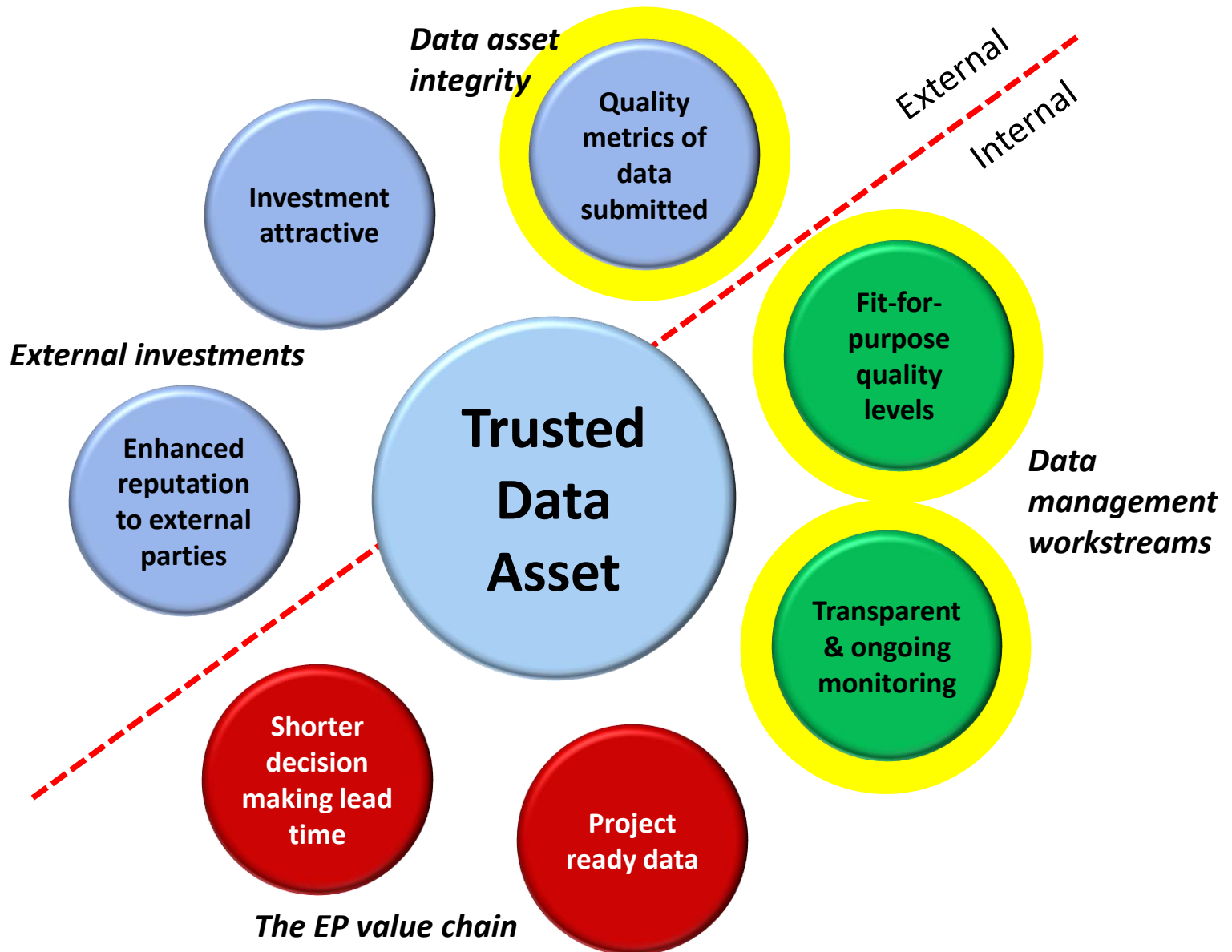
NDR11 Data Quality Workgroup Team Members

	Name
1	Helen Stephenson
2	Andrew Ochan
3	Marco Cota
4	Fanny Herawati
5	Melissa Amstelveen
6	Sarah Spinoccia
7	Jess Kozman
8	Ugur Algan
9	Richard Wylde
10	Cyril Dzreke
11	Lim TeckHuat
12	Hairel Dean
13	Deano Maling
14	Choo Chuan Heng
15	Iman Al-Farsi
16	Jill Lewis
17	Henri Blondelle
18	Armando Gomez
19	Kapil Joneja
20	Giuseppe Vitobello
21	Gareth Wright
22	Chan Kok Wah
23	Ali Alyahyaee (Scribe)
24	Philip Lesslar (Facilitator)

NDR2014 Data Quality Workgroup Team Members

2	TOOLS & DASHBOARDS
1	Kapil Jonjega
2	Jack Walten
3	Johanda du Toit
4	Gustavo Tinoco
5	Ferdinand Aniwa
6	David Atta-Peters
7	Angus Craig
8	Natalia Rakhmanina
9	Glab Khanuntin
10	Edem Mawuko
11	Alexander Kosolapov
12	Daniel Arthur
13	Eric Toogood
14	Tatiana Vassilieva
15	Henri Blondelle
16	Marianne Hansen
17	Jill Lewis
18	Mikhail Leypunsky
19	Aygun Mamedova
20	Rena Huseyn-zade
21	Irada Huseynova
22	Philip Lesslar

1	BUSINESS RULES
1	Helen Stephenson
2	Richard Salway
3	Abraham Oseng
4	Malcolm Flowers
5	Uffe Larsen
6	Calisto Nhatugues
7	Sylvester Nguessan
8	Gianluca Monachese
9	Jan Adolfssen
10	Lee Allison

3	DATA CORRECTION WORKFLOW
1	Mehman Yusufov
2	Vahid Jafarov
3	Aleksa Shchorlich
4	Ngwako Maguai
5	Joseph Justin Soosai
6	Samit Sencurta
7	Julian Pickering
8	Ugur Algan

What we did : 2014-2017

The National Data Repository
Data Quality Metrics Workstream

Part 1: Background and Case for Change

Prepared by:

The NDR Data Quality Working Group

Philip Lesslar
Helen Stephenson
Ugur Algan
Jill Lewis

energistics®
Energy Standards

The National Data Repository
Data Quality Metrics Workstream

Part 2: Business Rules Fundamentals

Prepared by:

The NDR Data Quality Working Group

Philip Lesslar
Helen Stephenson
Ugur Algan
Jill Lewis

energistics®
Energy Standards

The National Data Repository
Data Quality Metrics Workstream

Part 3 : Implementation

Prepared by:

The NDR Data Quality Working Group

Philip Lesslar
Helen Stephenson
Ugur Algan
Jill Lewis

energistics®
Energy Standards

Part 1: Background and Case for Change

Part 2: Business Rules Fundamentals

Part 3: Implementation

Context and justification to Management
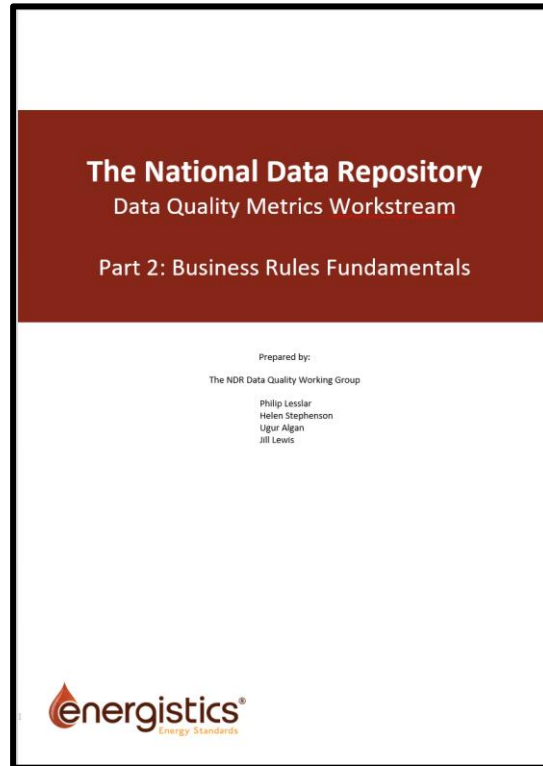
Data quality dimensions, key concepts around business rules, 18 data types, 241 rules
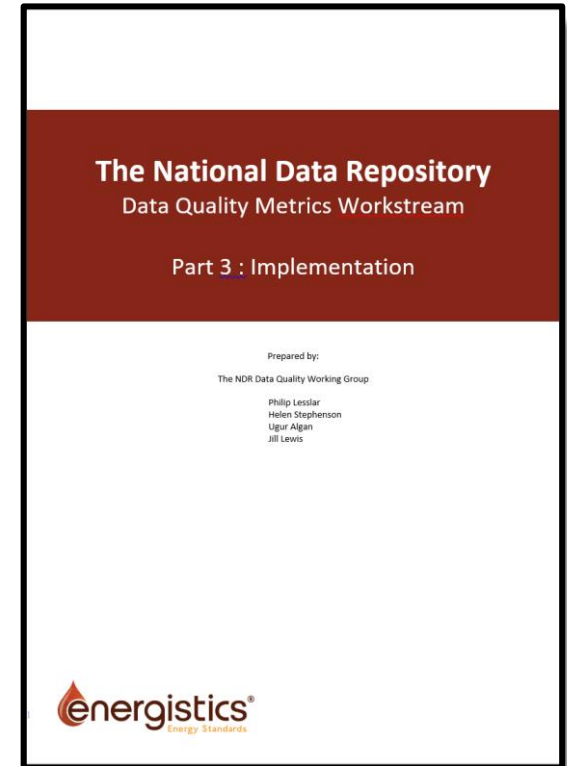
Metrics, dashboards, implementing rules as queries, understanding results, getting the program going

Why implement data quality metrics?

- Without metrics, we cannot measure the quality of the data we have
- Consequently, we cannot show how much quality, fit-for-purpose data there is...

Quality, fit-for-purpose data

Streamlines the business and its workflows

Increases data asset value and investor confidence

Builds essential data condition for effective use of new technologies

Perspective

Business

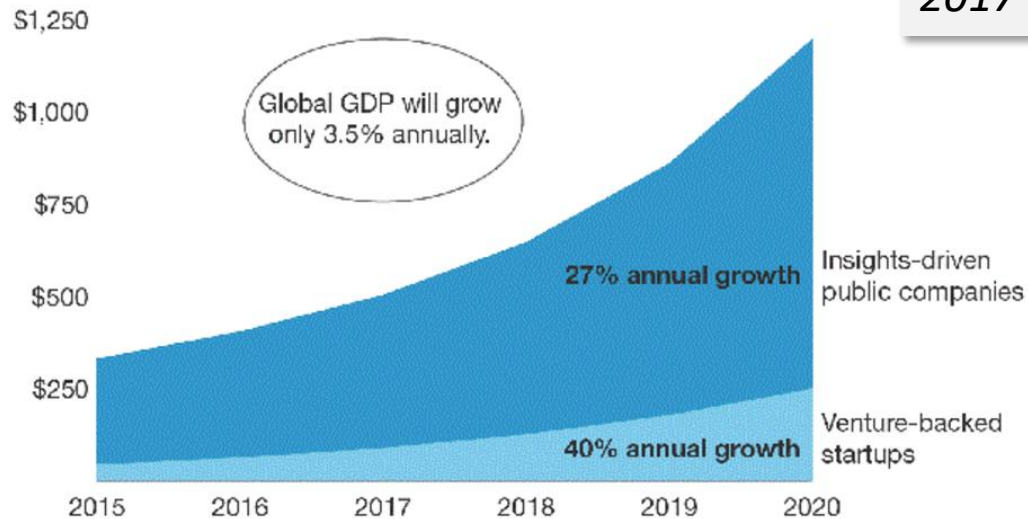Enabler for improving data efficiency by up to 90%

NDR

Data Management

Data Science & Analytics

Investment Trends

Insights-Driven Businesses Will Steal $1.2 Trillion Annually By 2020

Predictions 2017: Artificial Intelligence Will Drive The Insights Revolution

Revenue forecast of insights-driven businesses
($ billions)

Global GDP will grow only 3.5% annually.

27% annual growth — Insights-driven public companies

40% annual growth — Venture-backed startups

$1,250
$1,000
$750
$500
$250

2015 2016 2017 2018 2019 2020

Note: The data point for public companies in 2015 is actual revenue; all other data points shown are estimates or projected figures.

Source: Economic Intelligence Unit, Morningstar, and PitchBook Data

133325   Source: Forrester Research, Inc. Unauthorized reproduction, citation, or distribution prohibited.

Across all businesses, there will be *a greater than 300% increase in investment in artificial intelligence in 2017* compared with 2016.

Data Classification – Digital Data (>100 types in Upstream)

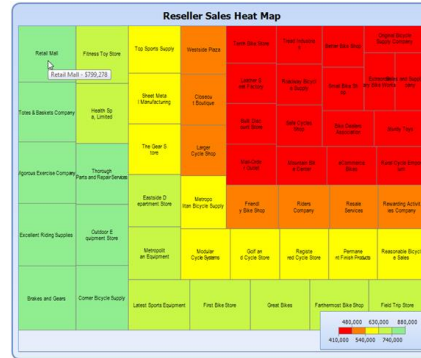| ← Primary Data →			← Secondary Data →	
Original Format Data	**Reference Data/ Metadata**	**Master Data/ Corporate "Single Source of Truth"**	**Derived Data**	**Data Collections**
Raw Seismic Raw Logs	Units of measure - *Linear measures* - *Pressure*	Static (hard) data - *Well header* - *Deviation* - *Checkshot* - *Temperature* - *Pressure*	Processed data - *Seismic deconvolution* - *Seismic filtering* - *Seismic processing* - *Edited logs* - *Spliced logs*	Composite data - *Completion log* - *Mud log* - *Paleontological composites* - *TRAPIS*
	Abbreviations - *TD, DFE, KB etc*	Interpreted (soft) data - *Geological markers* - *Seismic horizons*	Interpreted data - *Geological markers* - *Seismic horizons*	Data hoards - *Projects en masse* - *Personal stores* - *Team folders*
	Valid Lists			Data archive - *Projects en masse*
	Range indicators			
	Comments			
Requires: - *Official data repository*	Requires: - *Standards* - *Implementation across all impacted tools and databases*	Requires: - *Clear processes, workflows and checkpoints* - *Proper & official repository* - *Management and security processes around repository and data access*	Requires: - *Standard workflows* - *Standard algorithms* - *Standard processes* - *Housekeeping procedures*	Requires: - *Standard display and formatting templates* - *Procedures*

Open

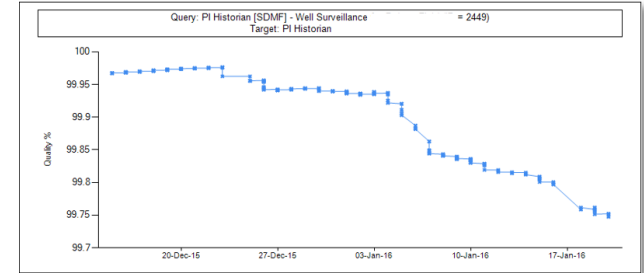# Packaging Quality Data – The Building Blocks

Quality Data Envelope

Data Type Building Blocks along the Exploration & Production (EP) Value Chain

PRIMARY DATA

SECONDARY DATA

Well header
Deviation
Checkshot
Pressure
Temperature

Plant Information (PI)

Geological markers
Geological zones
Seismic horizons

Exploration workflow

Facilities workflow

Production Geology workflow

Analytics workflow
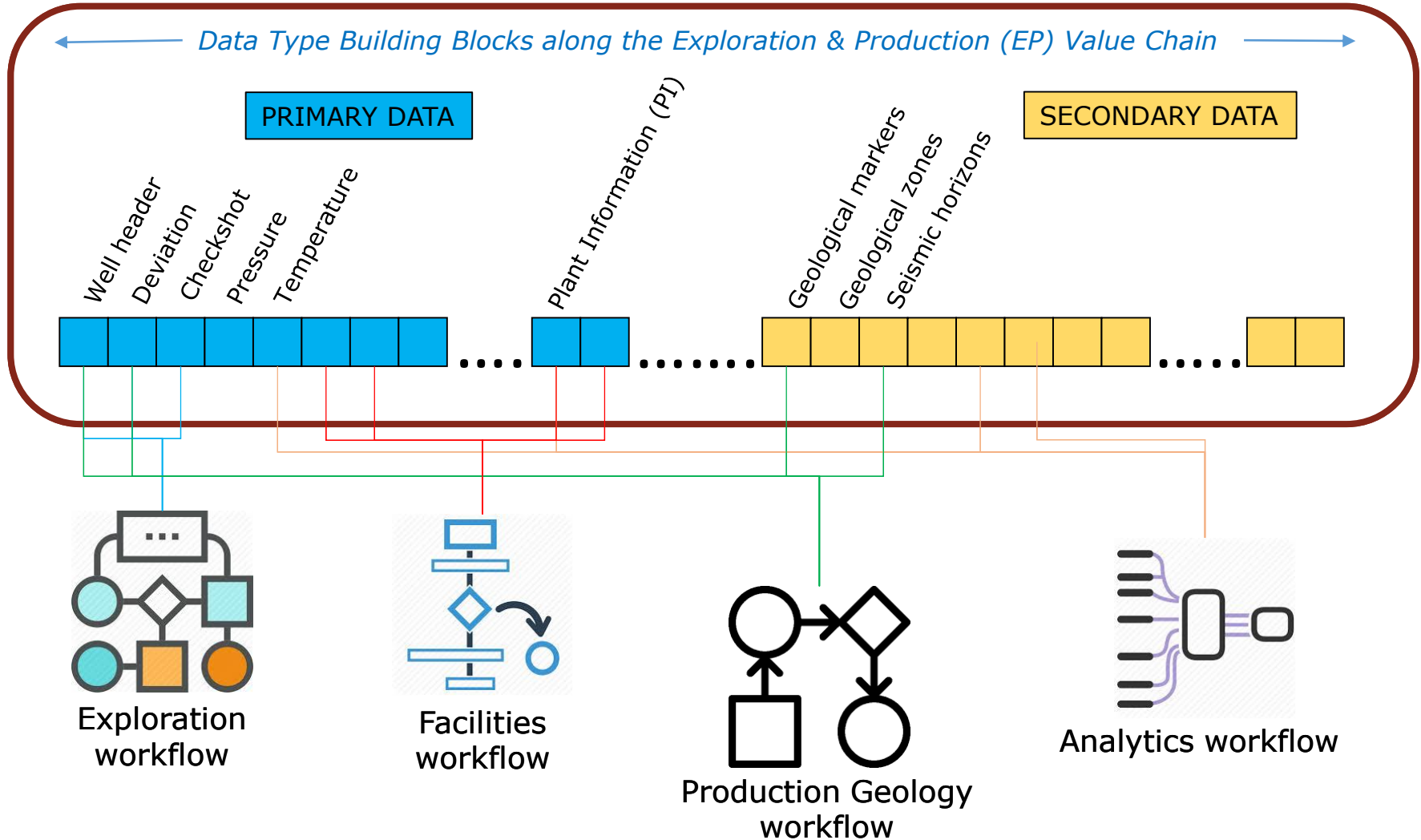
Data Science / Analytics – Typical Deliverables
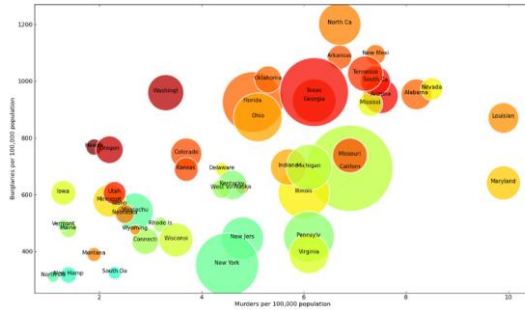
Bubble Plots

Heat Map

Trend Plot (IQM)
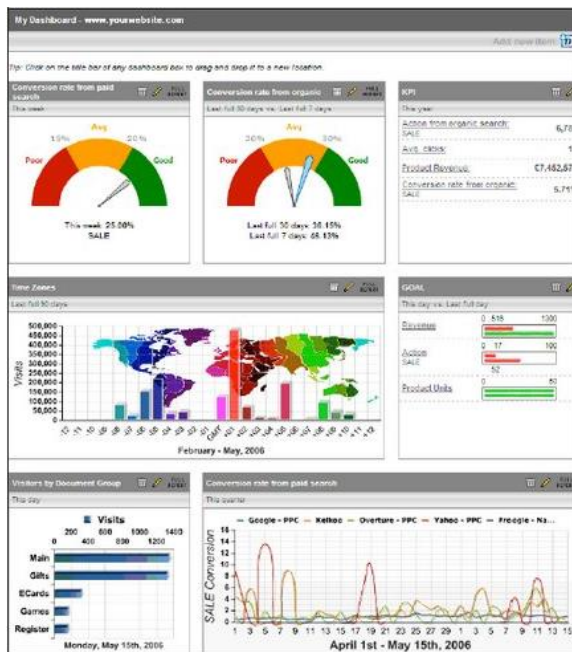
Yahoo Web Analytics
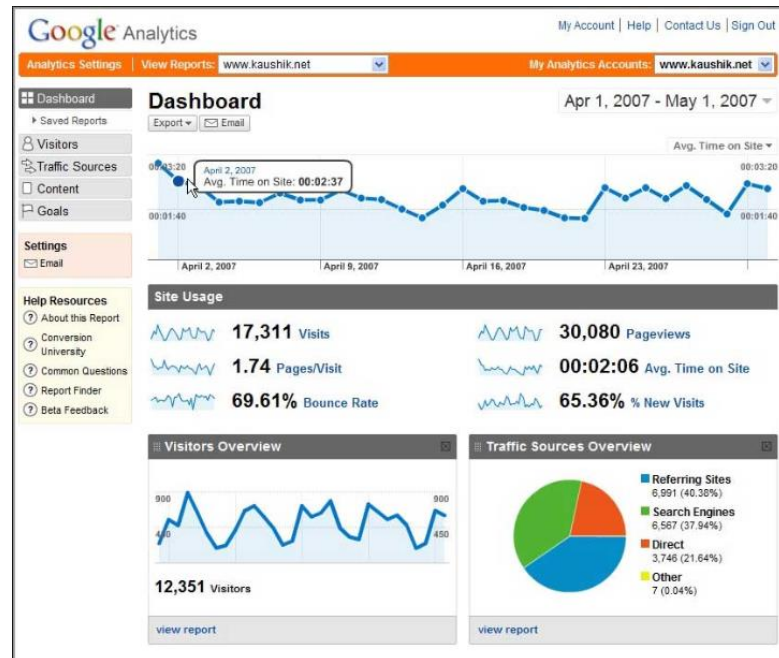
Google Analytics

Data Analytics Conceptual Architecture

Corporate
Databank
Stack

Data Extraction

-Mapping
-Extraction
-Cleansing
-Standardising
-Quality control

Requires quality data to be available

Right Decision

Data Viz & Analytics

Cleansed / Qced data does not flow back to the corporate banks

Timely Intervention

Note: The Data Mart starts to have better quality data than the official corporate databank
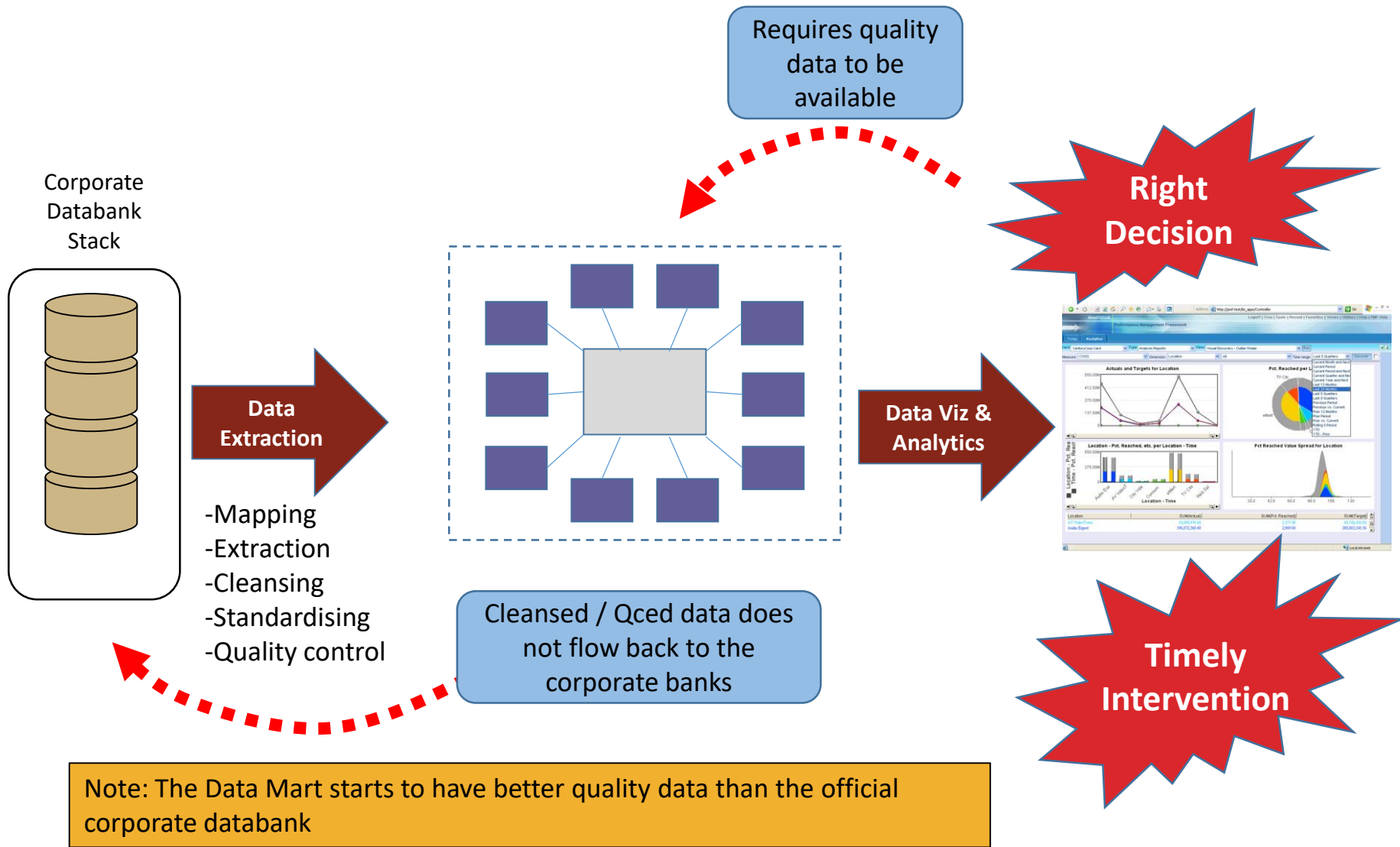
Data Quality Error Persistence

		Query ID	Query Name	Data Type	Quality Type	QD
→	◢	2448	SCE Test Data Approved By must be defined	Pressure Analysis	Completeness	74602.00 / 00.0 / 100.00 ─▲●
→	◢	2131	SCE Test Data must have Function Test defined	Pressure Analysis	Completeness	74602.00 / 10200.00 / 86.33 ▼▲▲
→	◢	2127	SCE Test Data must have Inflow Failure Mode defined	Pressure Analysis	Completeness	74602.00 / 3107.00 / 95.84 ▲▲▲
→	◢	2134	SCE Test Data Positive or Inflow Acceptable Leak Rate (psi/min) must be defined	Pressure Analysis	Completeness	74602.00 / 118.00 / 99.84
→	◢	2135	Well must have Annulus Pressure defined	Pressure Analysis	Completeness	27228.00 / 5176.00 / 80.99 ▼▲─
→	◢	2130	Well must have String Status defined	Pressure Analysis	Completeness	1385.00 / 94.00 / 93.21 ▲───
→	◢	2129	Well must have String Type defined	Pressure Analysis	Completeness	1385.00 / 00.0 / 100.00 ●●●

Business Rule:
Well must have annulus pressure defined

These errors will only be recognised if you are tracking the quality levels in the source databank

The analytics may not indicate quality levels

Eg. Annulus Pressure

Corporate Databank
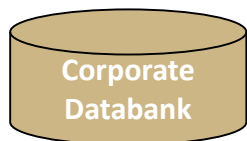
27228.00
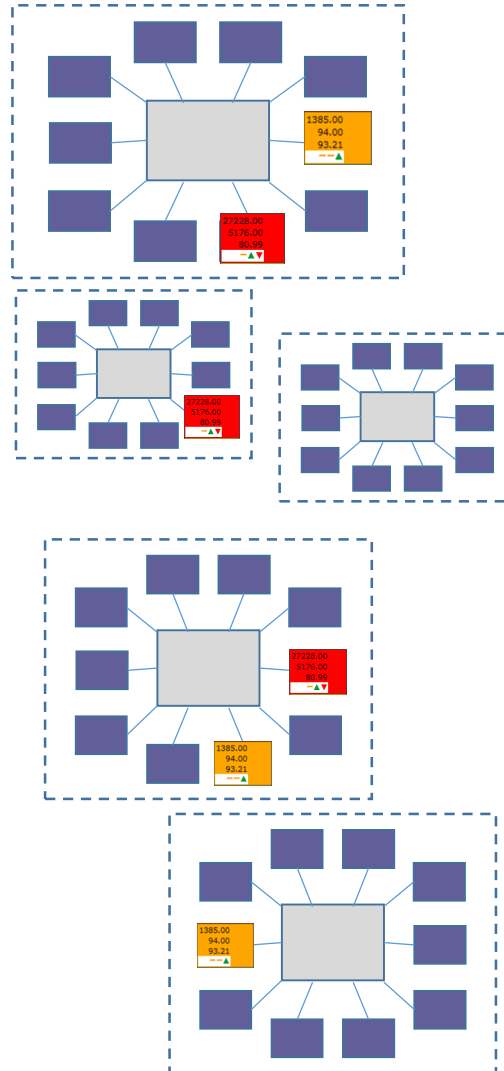5176.00
80.99
─▲▼
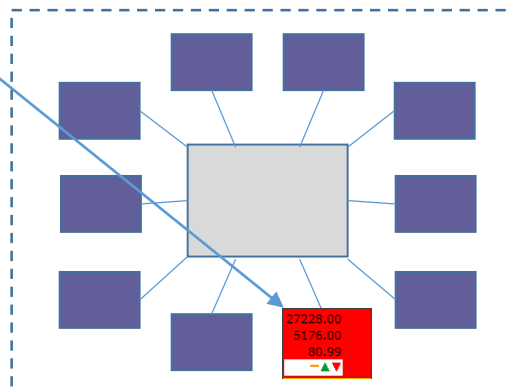
27228.00
5176.00
80.99
─▲▼
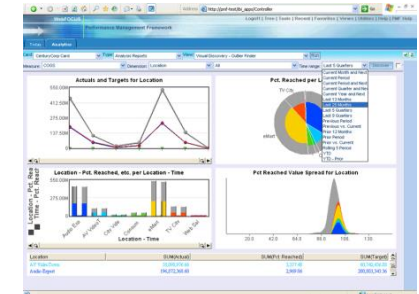
27228.00
5176.00
80.99
─▲▼

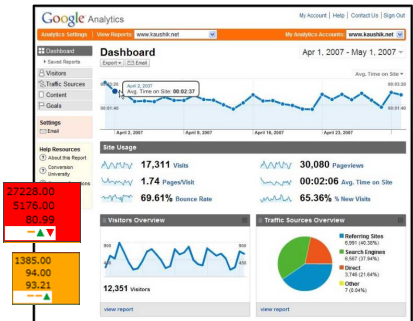# Data Quality – Progressive Lopsidedness + Hidden Risks

Poorer Quality
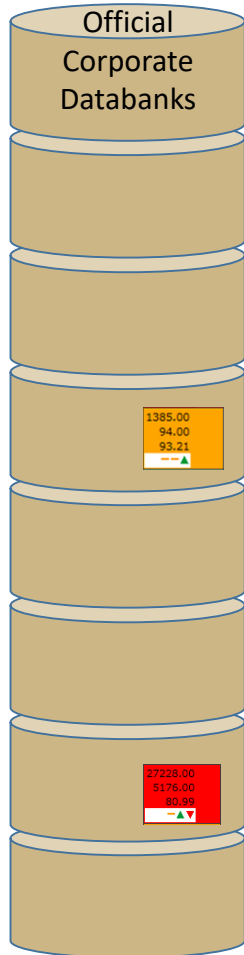
Better Quality

Right Decision?

Timely Intervention?

Official Corporate Databanks

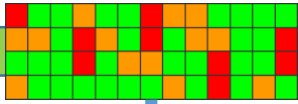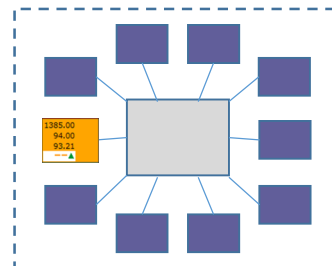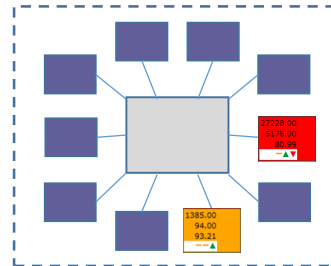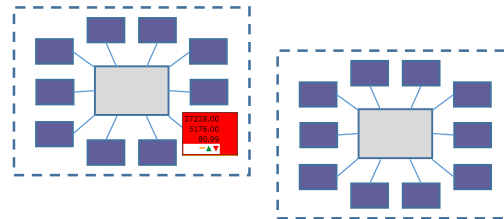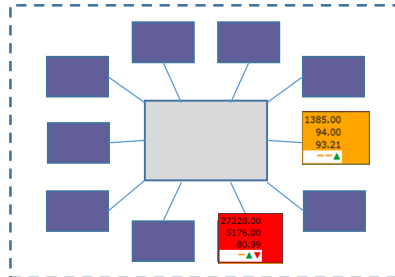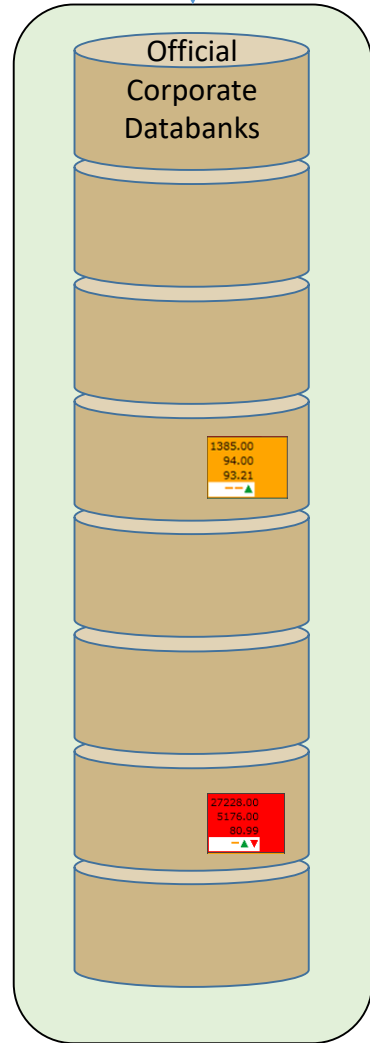# Data Quality Metrics – Tackling Quality at the Source
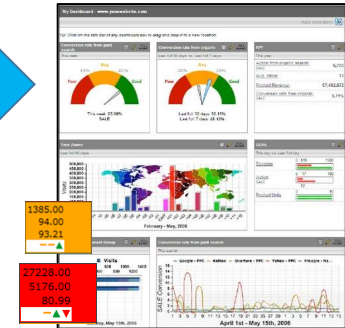

Data Quality Metrics Dashboard

Quality throughout the life cycle

Right Decision
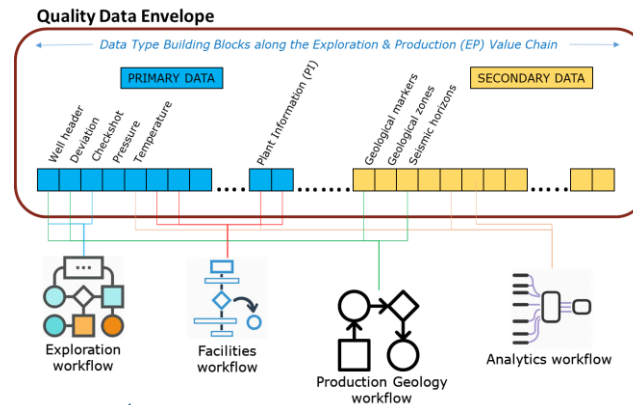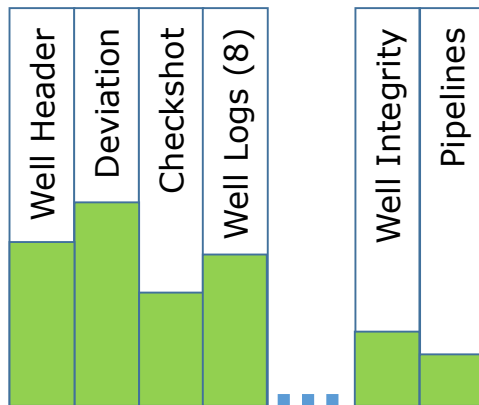
Timely Intervention

Official Corporate Databanks
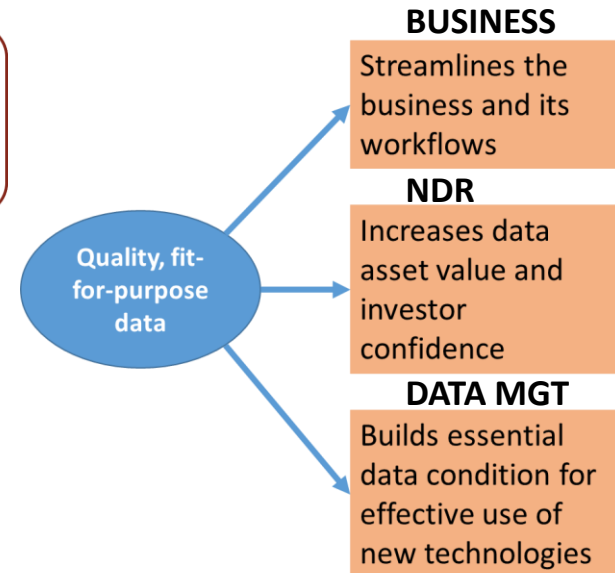
Concluding Remarks

- Understand our DATA INVENTORY
- Implement METRICS to improve QUALITY

- Address data types as building blocks across all 100+ EP types

- We solve business problems and create new opportunities
- While measuring and knowing where we are at all times

- Measure and KNOW how much FIT-FOR-PURPOSE data there is

BUSINESS
Streamlines the business and its workflows

NDR
Increases data asset value and investor confidence

DATA MGT
Builds essential data condition for effective use of new technologies

Towards data science and big data analytics, by putting science into data management

Thank You