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



NATIONAL DATA REPOSITORY

STAVANGER NORWAY

Sponsored by:



Organized by:



- 165 participants
- **Representing 30 countries**

, 6-8 June, Stavanger, Norway



Google Maps

Data Quality Metrics – The Journey



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	The National Data Repository Data Quality Metrics Workstream Part 2: Business Rules Fundamentals	The National Data Repository Data Quality Metrics Workstream Part 3 : Implementation	
Prepared by: The NDR Data Quality Working Group Philip LesiSar Helen Stephenson Ugu Algan Jill Lewis	Prepared by: The NOR Data Cuality Working Group Philip Lesslar Helen Stephenson Ugur Algan Jill Lewis	Prepared by: The NDR Data Cuality Working Group Philip Lessiar Helen Stephenson Ugur Algan Jill Lewis	
Part 1: Background and	Part 2: Business Rules	Part 3: Implementation	
Case for Change	Fundamentals	Metrics, dashboards,	
	Data quality dimensions,	implementing rules as	
Context and justification	key concepts around business rules 18 data	queries, understanding	
to Management	types, 241 rules	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, fitfor-purpose data there is...

Quality, fitfor-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

FORRESTER' RESEARCH

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

Predictions 2017: Artificial Intelligence Will Drive The Insights Revolution



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

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

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 - <i>Linear measures</i> - <i>Pressure</i>	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 - <i>Projects en masse</i>	
	Range indicators				
	Comments				
Requires: - <i>Official data</i> <i>repository</i>	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	

Packaging Quality Data – The Building Blocks



Data Science / Analytics – Typical Deliverables







Heat Map



Yahoo Web Analytics



Google Analytics

Data Analytics Conceptual Architecture



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

Data Quality Error Persistence





-47

Data Quality – Progressive Lopsidedness + Hidden Risks



Data Quality Metrics – Tackling Quality at the Source



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



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

Thank You