



**PETRONAS**

# Data Quality Metrics – Considerations for Achieving Trusted Data

**Philip Lesslar**

Technical Assurance, Group Technical Data, PETRONAS

**Digital Energy Conference**

5<sup>th</sup> October 2016

Impiana Hotel, Kuala Lumpur

© 2014 PETROLIAM NASIONAL BERHAD (PETRONAS)

All rights reserved. No part of this document may be reproduced, stored in a retrieval system or transmitted in any form or by any means (electronic, mechanical, photocopying, recording or otherwise) without the permission of the copyright owner.

# Contents

- Trustworthiness and the need for Trusted Data
- Data Quality Metrics (DQM)
- How far can we go with DQM? – The Expert Factor
- So how do we get to trusted data?
- The Quality Data Inventory
- Packaging quality data – The building blocks
- Data delivery and Data Access and Retrieval (DART)
- Quantified benefits and sustainability
- Summary

# Trustworthiness and the Need for Trusted Data

## Problem Statement

1. Data collection consumes about half the time in business activities.
  - 52% in 10 workflows studied.
2. Everyone who acquires data spends time re-checking it.
  - 58 man years/year for well header, check shot, deviation & basic logs (find & quality check).
3. Data come from various sources.

# Our 5-step Approach Towards Trusted Data...



- Single source of truth for each data type
- Identified mandatory and optional attributes needed for business activities
- Standard and approved process for quality checking (QC) of the data
- Identified Data Custodians accountable for ensuring data capture, availability and longevity
- Business appointed Subject Matter Expert (SME) to advise and QC
- Subject Matter Focal (SMF)
- Audit trail & documentation

# Data Quality Metrics (DQM) Simplified

It is about measuring data quality...

- 1 Business rules ↪ E.g. well header must have total depth defined.
- 2 Encoded metrics in application ↪ Usually written as sequel (SQL) queries or internal proprietary syntax.
- 3 Data Quality Dashboard

Query Set Name	Type	Lower hours	Upper hours	A	B	C	D	E	F	G	H
Checkshots / Time-Depth	datatype	90.0	95.0	38923.0	141485.0	219232.0	179813.0	547160.0	8083.0	199417.0	447303.0
				1342.0	1059.0	14651.0	1649.0	54092.0	726.0	8854.0	11968.0
				96.6	99.3	93.8	99.1	90.1	91.0	93.6	97.3
Formation Tops	datatype	90.0	95.0	608295.0	936564.0	4069835.0	2521585.0	4470170.0	71720.0	1198380.0	5571120.0
				23955.0	10101.0	91639.0	46544.0	22126.0	2915.0	13638.0	92125.0
				96.1	98.9	97.7	98.2	99.5	95.9	98.9	90.3
Position Logs	datatype	90.0	95.0	43608.0	47375.0	164842.0	79362.0	382958.0	3978.0	545683.0	249982.0
				95.0	2418.0	3648.0	992.0	62900.0	104.0	1712.0	5747.0
				99.8	94.8	97.8	98.8	83.4	97.4	99.7	97.7
Seismic Horizons Faults	datatype	90.0	95.0	297620.0	538844.0	1810189.0	1110480.0	886556.0	7212.0	117806.0	1221882.0
				19072.0	18895.0	140733.0	252875.0	52738.0	489.0	2800.0	68455.0
				93.4	96.5	94.1	93.8	93.5	93.2	97.6	94.4
Well Headers	datatype	90.0	95.0	43452.0	81288.0	238046.0	138744.0	716652.0	7482.0	789508.0	873184.0
				1525.0	3638.0	9922.0	13816.0	47282.0	350.0	17838.0	83964.0
				96.6	95.5	95.8	90.0	93.4	95.3	97.7	90.4

User friendly traffic light dashboard allowing for intervention actions.

\*IQM screenshot (courtesy of Exprodat Inc.)

- Features:
- Automated rule checking
  - Consistent
  - Progressive
  - Targeted



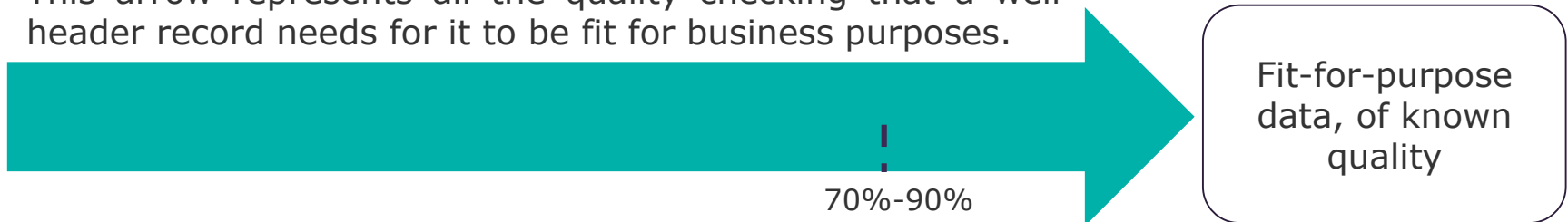
Total records polled  
 Total errors found  
 Quality percent  
 Trends : Monthly, Weekly, Daily  
 $(1 - (1059 / 141485)) * 100 = 99.25$



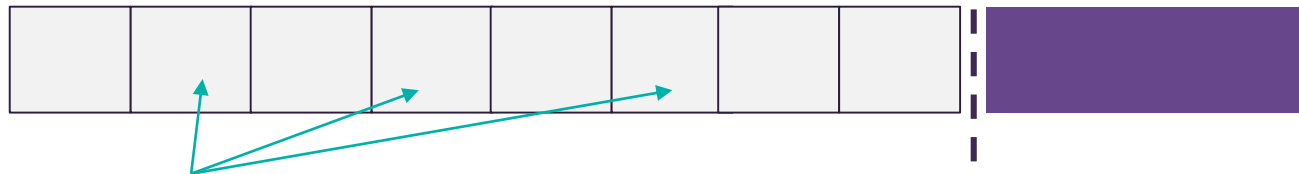
# How Far Can We Go with Data Quality Metrics?

Given any data type (e.g. well header)...

This arrow represents all the quality checking that a well header record needs for it to be fit for business purposes.



70%-90%



## Business rules

Checks:

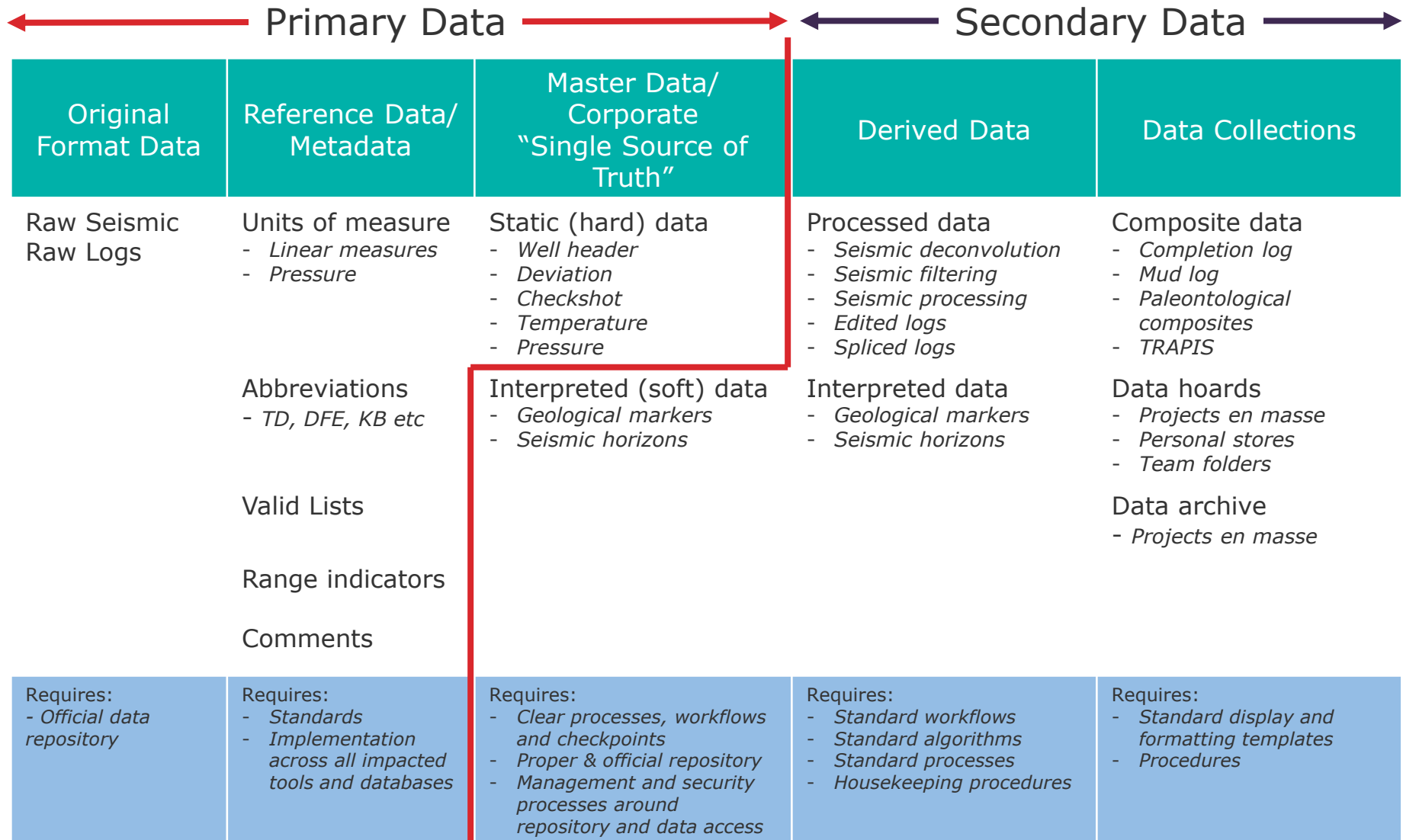
- Mandatory attributes
- Data quality dimensions – completeness, duplication, conformity, consistency, integrity, accuracy

## Subject Matter Expert

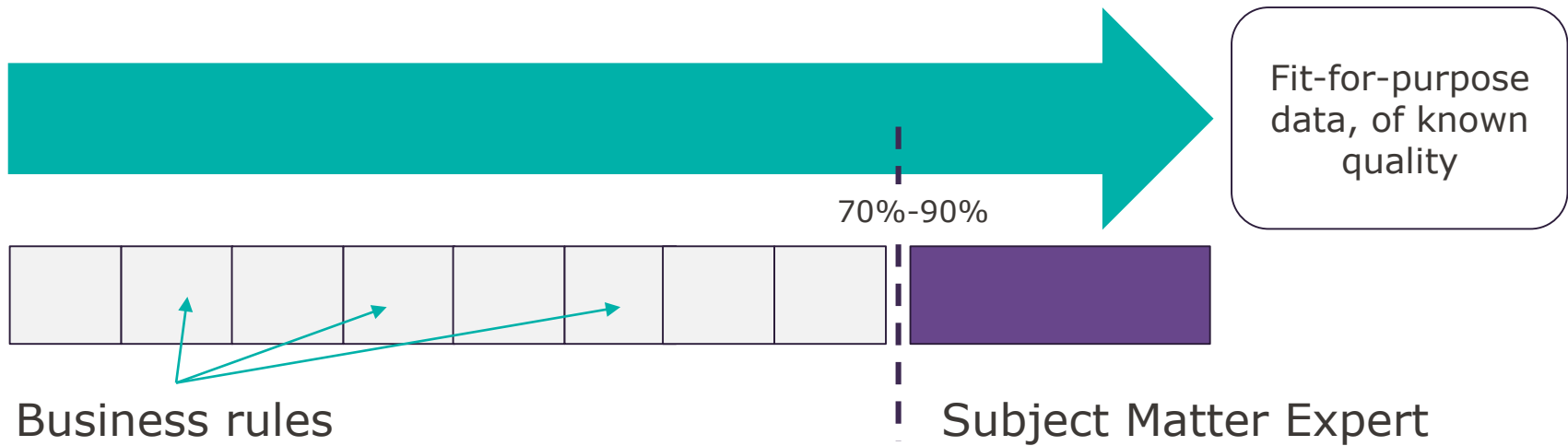
Checks:

- Relevant control documents
- Significance of missing values
- Referential integrity
- QC & assurance
- Fit for purpose
- Assignment of quality indicator

# Data Classification – Digital Data



# So How Do We Get To Trusted Data?



**DQM**  
**Data Quality Metrics**

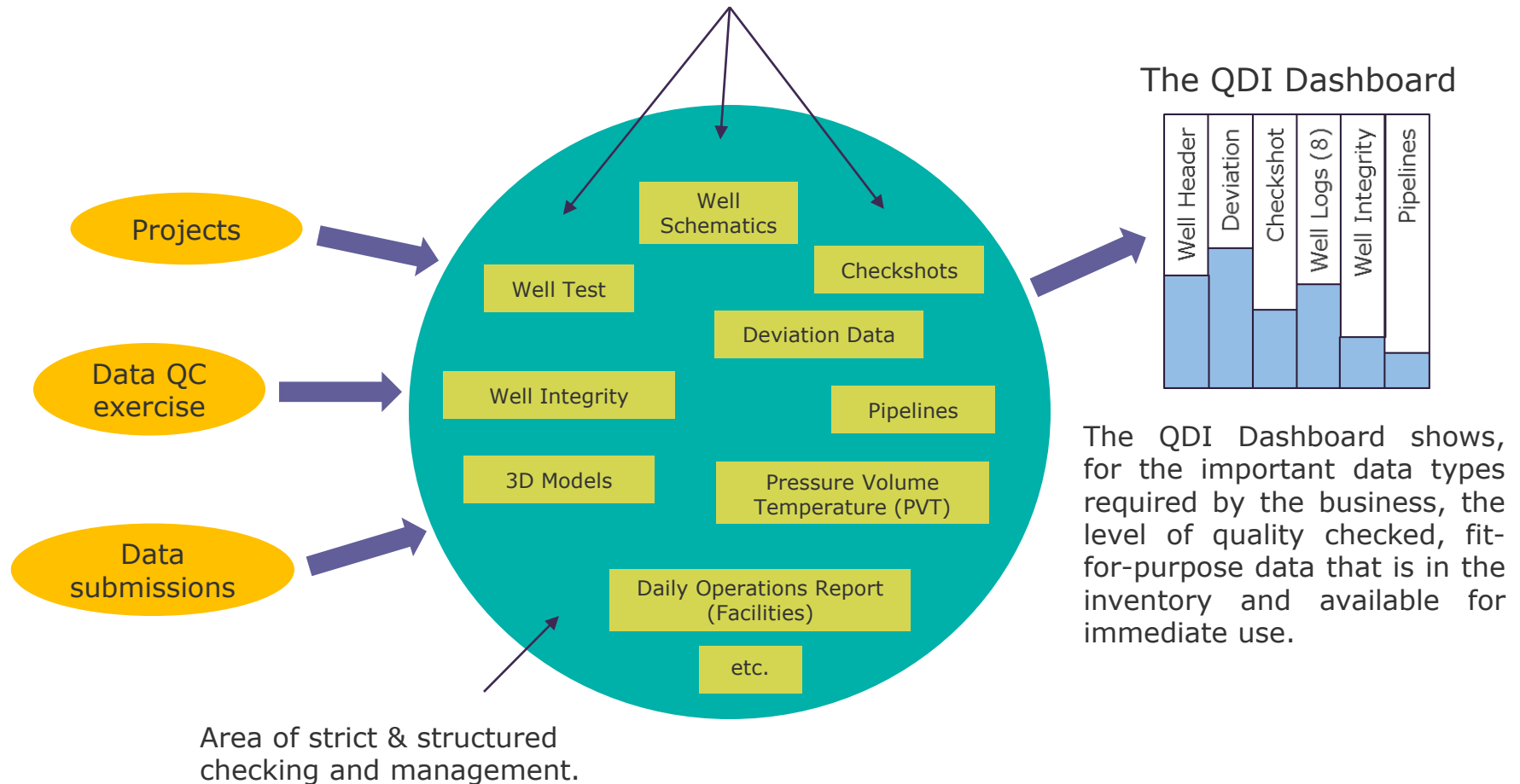


**QDI**  
**Quality Data Inventory**



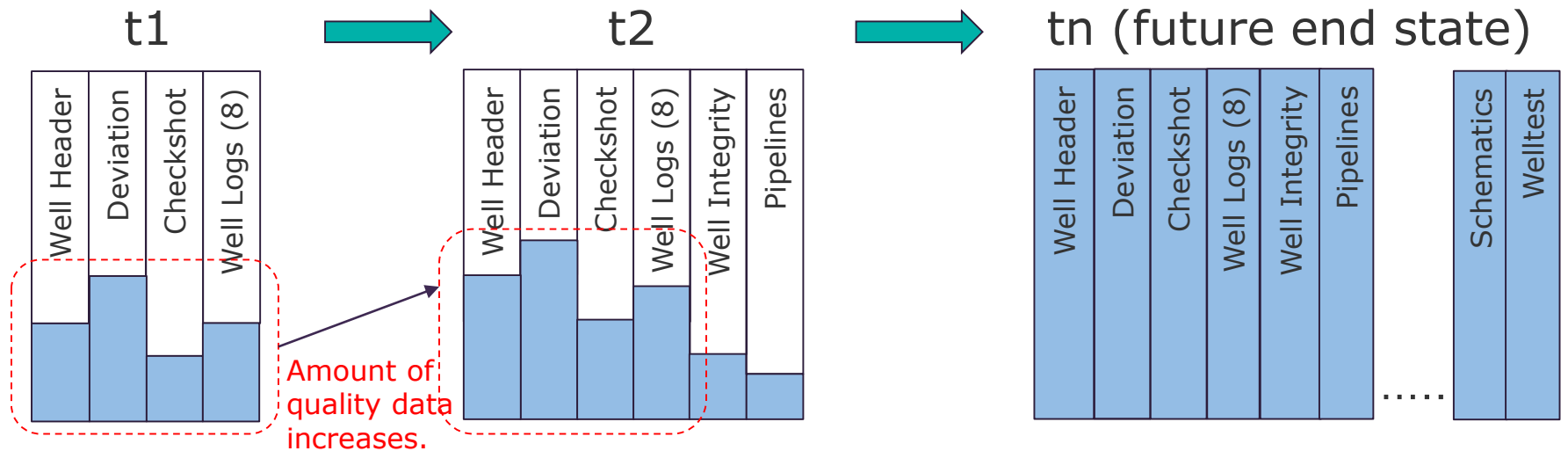
# What is the Quality Data Inventory (QDI)?

Collective of fit-for-purpose quality checked data, required for business activities, and that can be re-used in the company.



# What is the Purpose of the Quality Data Inventory?

To provide a structured, consistent and transparent way of showing and demonstrating progress in addressing quality requirements in our data.



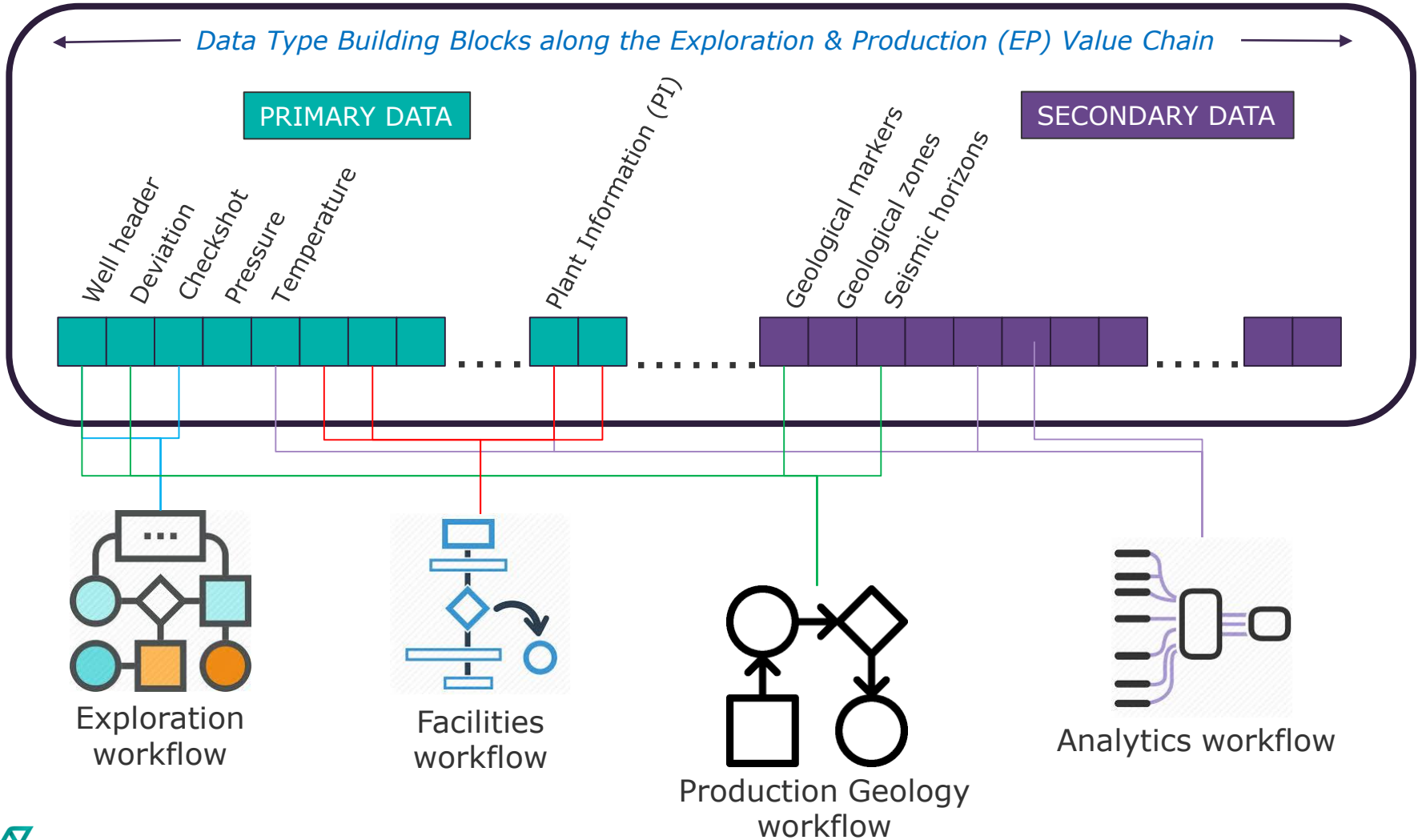
Over time, the data gets better and better, progress can be measured and targets can be set.

All important data that is required for the business is properly checked, with signoffs by appropriate authority.

## Filling up the tanks with Trusted Data

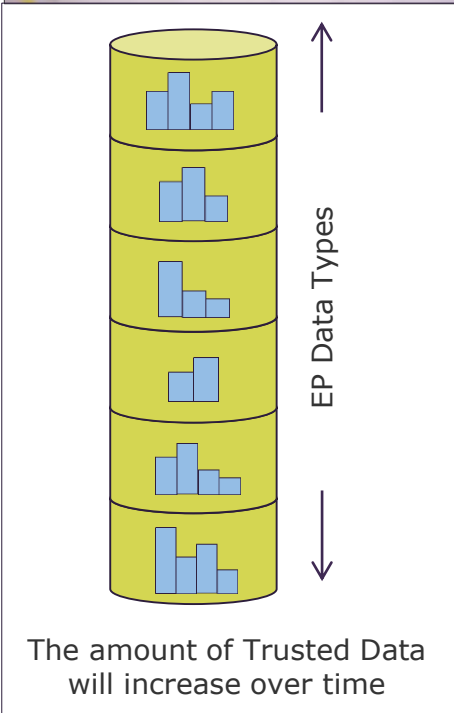
# Packaging Quality Data - The Building Blocks

## Quality Data Envelope

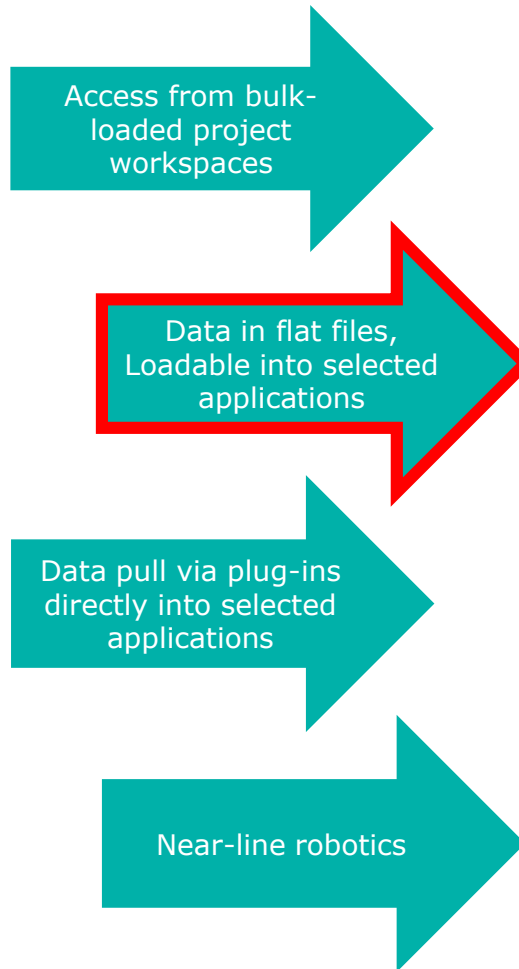


# Data Delivery Methods

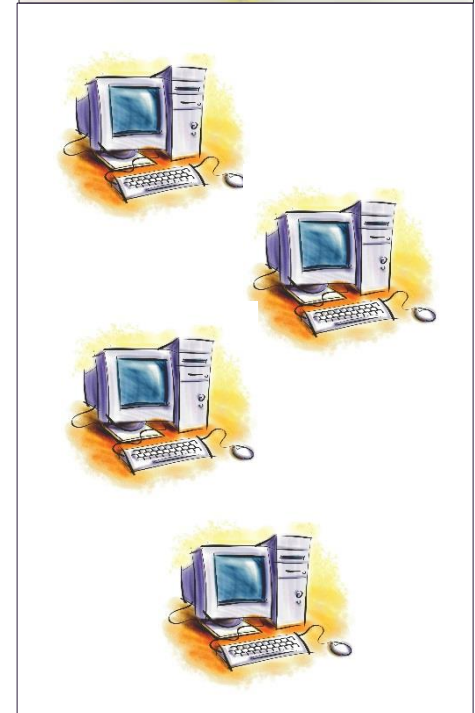
## Official Databanks



## Data Delivery Mechanisms



## The User Environment



# Data Delivery – The DART project

Welcome Philip Lesslar (GTD/PD&T)

## DATA QUALITY METRICS DASHBOARD

PDW Repository | Dashboard | Wells | QDI Quality | QDI Completeness | DART | Administration | IQM

### Data Access & Retrieval

Well Name :

\* Please enter the Well name to search

File Type

\* File type (CSV : For easy viewing of data & TAB : For loading to petrel)

Total Records : 194

UWI	WELL NAME	FIELD NAME	COUNTRY NAME	DataType1	QCed	DataType2	QCed	DataType3	QCed	DataType4	QCed	DataType5	QCed
10				<a href="#">WELL HEADER</a>	NO	DEVIATION(0)	NO	CHECKSHOT(0)	NO	PICK(0)	NO	STRAT(0)	NO
10				<a href="#">WELL HEADER</a>	NO	DEVIATION(0)	NO	CHECKSHOT(0)	NO	PICK(0)	NO	STRAT(0)	NO
10				<a href="#">WELL HEADER</a>	YES	DEVIATION(0)	NO	<a href="#">CHECKSHOT(16)</a>	YES				
10				<a href="#">WELL HEADER</a>	NO	DEVIATION(0)	NO	CHECKSHOT(0)	NO				
10				<a href="#">WELL HEADER</a>	YES	<a href="#">DEVIATION(27)</a>	NO	<a href="#">CHECKSHOT(18)</a>	YES				
10				<a href="#">WELL HEADER</a>	YES	<a href="#">DEVIATION(13)</a>	NO	<a href="#">CHECKSHOT(24)</a>	YES				
10				<a href="#">WELL HEADER</a>	YES	DEVIATION(0)	NO	CHECKSHOT(0)	NO				
10				<a href="#">WELL HEADER</a>	YES	DEVIATION(0)	NO	CHECKSHOT(0)	NO				
10				<a href="#">WELL HEADER</a>	NO	<a href="#">DEVIATION(90)</a>	NO	CHECKSHOT(0)	NO				

WELL INFO

```

#22 GEOGRAPHIC CRS : World Geodetic System 1984
#20 PROJECTED CRS : WGS 84 / UTM zone 50R
#RIG ELEVATION DATUM : KB
#RIG ELEVATION ABOVE UCD : 47
#VERTICAL REFERENCE UNITS OF MEASURE : feet
#HORIZ REFERENCE : Geoid
=====
MEASURED DEPTH(meters)    INCLINATION(degs)    AZIMUTH(degs)
TRUE VERT DEPTH(meters)
=====
104.55    1.00    -999.00    -999.00
329.13    0.75    -999.00    -999.00
732.87    0.90    -999.00    -999.00
1078.99    0.75    -999.00    -999.00
1142.70    0.75    -999.00    -999.00
1583.82    1.00    -999.00    -999.00
1589.26    1.50    -999.00    -999.00
1906.69    1.75    -999.00    -999.00
2137.24    1.00    -999.00    -999.00
2446.32    1.25    -999.00    -999.00
2999.94    1.75    -999.00    -999.00
2650.42    1.00    -999.00    -999.00
2720.98    0.75    -999.00    -999.00
2735.58    1.00    304.00    2735.58
2742.29    1.75    289.00    2742.29
2765.39    4.50    278.00    2765.39
2762.40    5.00    273.00    2762.40
2779.78    6.50    269.00    2779.47
2815.69    9.00    265.00    2815.18
2866.03    9.50    261.00    2865.12
2933.09    10.00    259.00    2930.96
2960.09    10.00    262.00    2964.12
3037.75    20.00    281.00    3050.44
3115.06    20.25    282.00    3104.39
3192.48    20.25    281.00    3177.24
3283.69    18.50    280.00    3267.74
3352.80    18.50    280.00    3331.16
                    
```

```

#WELL INFO#
#
#
#
#
#GEOGRAPHIC COORDINATE CRS : World Geodetic System 1984
#RIG_ELEVATION : 47(feet)
#RIG_ELEVATION DATUM : KB
#DATUM REFERENCE : MSL
#DEPTH MODE : TVDBDD
#TIME TYPE : two-way
#
#IMPORTANT INFO: FOR OFFSHORE WELL, TRUE VERTICAL DEPTH BELOW
TIME DEPTH DATUM(TVDBDD) IS MSL/SRD/SS, IF ONSHORE TVDBDD IS
GL #
#
#
#DEPTH(meters) (DepthMode=TVDBDD) TIME(milliseconds)
(TimeType=two-way)
=====
564.79    570
842.16    786
961.64    874
1087.22    962
1189.02    1030
1340.21    1130
1460.91    1210
1590.14    1290
1687.98    1350
1808.38    1422
1875.43    1462
2010.16    1538
2152.80    1618
2238.76    1664
2284.48    1688
2396.64    1748
2555.14    1832
2680.11    1894
                    
```



# Quantified Benefits And Sustainability

<b>What problems are we trying to solve?</b>	<b>Develop &amp; implement the solutions</b>	<b>Measure how much of the problem has been solved</b>	<b>Implement continuous tracking to move the dot</b>
<ul style="list-style-type: none"><li>• Quantify problems:<ul style="list-style-type: none"><li>– 58 manhrs/yr NPT.</li><li>– 52% data downtime.</li></ul></li><li>• Identify causal factors:<ul style="list-style-type: none"><li>– No framework.</li><li>– Fragmented data.</li><li>– Unclear inventory.</li><li>– Unknown quality.</li><li>– Access difficulties.</li><li>– Long delivery time.</li></ul></li></ul>	<ul style="list-style-type: none"><li>• Prioritize key issues.</li><li>• Develop framework and end-to-end working model across key data types.</li><li>• Deliver data as early as possible for clarity and transparency.</li><li>• Secure additional commitments.</li></ul>	<ul style="list-style-type: none"><li>• Quantification methodology.</li><li>• Ensure capture of initial states in quantified terms.</li><li>• Start calculating deltas as soon as solution in place.</li></ul>	<ul style="list-style-type: none"><li>• Track continuous progress via register.</li><li>• Ensure broad awareness of progress for continued momentum.</li></ul>

# Summary

- Data Quality Metrics go a long way in improving data quality but not to trusted level.
  - We have our Data Quality Metrics (DQM) programme and are running daily ~1000 queries against about 28 database instances.
  - We also have Quality Data Inventory (QDI) which combines DQM with SME checks to bring data to trusted levels. This is tracked via an automated dashboard.
- A database filled with quality data is of no value until it is easily accessible.
  - Our Data Access and ReTrieval (DART) project allows users to find and retrieve data quickly.
- We need to track progress and value of quality data.
  - Level of quality data is automatically tracked in dashboard for hands-free transparency
  - We have implemented a methodology to calculate and track quantified benefits.
- Sustainability is a function of good processes, transparency, quality and accessibility.



**PETRONAS**

**Thank you**