Digital technologies, DIGITALISATION and the importance of data quality

Duncan Irving

Finding Petroleum, November 2017
Digitalisation? ...not another buzzword, please

“It is wrong to suppose that if you can’t measure it, you can’t manage it – a costly myth.”

W Edwards Deming
But didn’t we do the Digital Oilfield a decade ago?

“No, the Digital Oilfield is a bunch of operational technologies that operate in an effectively closed loop.

They never linked the oilfield to the supply chain depots or finance and planning functions.

…or even operations to the subsurface”
Integrating HSSE data across other business functions for wider impact

H2S (hydrogen sulphide) can be deadly even at low concentrations. Wells are properly equipped for safety monitoring. Sulphide deposits also impact flow rates, and subsequent corrosion compromises asset integrity. Data was stored in HSE, ops and maintenance silos.

The drive for improved safety led to new service offerings when integrated with operational, resourcing and planning data:

- More competitive pricing structures
- Improved revenue collection
- Better equipment utilization
- Millions saved in payroll audit / time keeping process
- Analysis provides impact of having added 141 locomotives to fleet

“Estimated benefits from the utilization of Union Pacific’s enterprise data warehouse is in excess of $100 Million each year.”

(Union Pacific)

Analysis of track side data collected for safety improvements:

- Bearing-related derailments reduced by 75% - hundreds of minor incidents per year, plus some major ones
- 20 million daily pattern matches of sensor spot 1500 issues/day with problem wheels

Infrared sensors every 30 km gather wheel temperature and send to operations centre.

3350 trains per day (hundreds of 150 yr. old axles each) 45,000 km of track

The drive for improved safety led to new service offerings when integrated with operational, resourcing and planning data:

- Understanding of how to mitigate sulphide deposits
- More effective well planning and production engineering
- Ensure correct equipment and people are assigned to drilling and interventions based on H2S shows

Strong safety regime was paramount – but more well plans completed on time, and higher production assurance achieved as a result of real-time data integration.

Estimation from the utilization of Union Pacific’s enterprise data warehouse is in excess of $100 Million each year.”

(Union Pacific)
Data-driven maintenance powers business-critical capabilities

- “Performance-based maintenance” make Siemens the operator of the most reliable high-speed train on RENFE network
- One in 2,300 rides noticeably delayed: key criterion for business success: passengers reimbursed when delay >15 mins

  - Sensors measure constantly key parameters e.g. traction motor bearings
  - Analytics enables a stable incident prediction based on strong understanding of patterns
  - Abnormal patterns trigger inspection to prevent failure on track

Piston ring failures are infrequent events (and hence little data), but can immobilize a vessel at sea. Data integration exposed statistical relationships in behavior that was not seen in conventional data silos.

  - Delivered >10-day lead time on predictions for component failure
  - 75% of failures (target: 60-80%) captured by sensor-based analytics
  - 63% accuracy in prediction (target: 50-65%) of piston condition

This new understanding of failure of fundamental components such as piston rings could be acted on for productionised business impact.
Five Key Observations from Upstream studies

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Organisational constructs, culture and the <strong>project mentality</strong> make project delivery difficult</td>
</tr>
<tr>
<td>2</td>
<td><strong>Business Requirements</strong> gathering and prioritisation needs improving</td>
</tr>
<tr>
<td>3</td>
<td>Key common <strong>Data Management capabilities</strong> not in place across upstream</td>
</tr>
<tr>
<td>4</td>
<td><strong>Definitions</strong> – key terms mean different things to different people – means communication fraught with misunderstanding</td>
</tr>
<tr>
<td>5</td>
<td><strong>Future Data Platform</strong> IT architecture often <strong>lacks integration capabilities</strong></td>
</tr>
</tbody>
</table>
Organisational and Cultural Challenges

Issues arising from a project-centric approach

**Funding**
- Parochial thinking
- Departmental funding leads to departmental solutions

**Planning**
- Lack of strategic planning and foresight
- Missed opportunities to share data resources

**Coordination**
- Many projects with limited sharing and coordination
- Projects designed to meet specific business objectives

**Delivery**
- Project mentality gets in the way of delivering consistent architecture

**Maintenance**
- Lack of a "maintain" strategy: "Projects are dead as soon as they've been delivered"

**User Experience**
- No interoperability
- Suboptimal business impact

Governance is a key driver to a coherent data strategy but is not addressed strategically
Assemble an analytics team

If you care about becoming a Digital Upstream company you need to nurture these capabilities.

- **DELIVERY EXCELLENCE**
  Focus on Business Value and drive toward success. A technical expert, hands-on project and team manager.

- **DATA ENGINEER**
  Ability to acquire and wrangle data in multiple formats, including data transformations and integration. A software engineer with strong knowledge of best practices in application development.

- **INFRASTRUCTURE**
  A hands-on expert administrator of the production platform to be created and employed, they will optimise the solution and administrate the production system.

- **BUSINESS EXPERT**
  A domain expert adds contextual mapping, internal resourcing, and understanding of existing business processes.

- **DATA EXPERT**
  A data expert provides domain context around the metadata, master data and measurement data, and understands how and why it is generated and used.

- **DATA SCIENTIST**
  Cross industry experience in applying advanced analytics to a range of data types to generate valuable insights.

© Teradata 2017
What do these people do for me?

Data Science or "Data R&D" is concerned with the *discovery* of hidden patterns and insights.

“...data mining is an *exploratory* undertaking closer to research and development than it is to engineering. The CRISP cycle is based around *exploration*; it iterates on approaches and strategy... outcomes are far less certain, *and the results of a given step may change the fundamental understanding of the*...”

*Data Science For Business, Foster Provost & Tom Fawcet, my emphases*
Multi-genre Analytics Evolution

1. **REPORTING**
   - What is happening?

2. **QUERY & DRILL DOWN**
   - What exactly is the problem?

3. **ALERTS**
   - What actions are needed?

4. **FORECASTING**
   - What are the hidden patterns? What will happen next?

5. **SIMULATION**
   - What could happen?

6. **MACHINE LEARNING**
   - Self-learning systems with regression.

7. **DEEP LEARNING**

**Timeline:**
- **1990s**
- **2000s**
- **2010s**
But analytics have impact if we apply common dimensions of **Data Quality**

Quality should be consistently defined and monitored to drive performance measuring

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Description</th>
<th>Conformance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accessibility</td>
<td>A measure of information access / availability</td>
<td>Data is available for all users</td>
</tr>
<tr>
<td>Accuracy</td>
<td>A measure of information correctness</td>
<td>A flow rate of $12.7 \text{ m}^3\text{s}^{-1}$ really is stored as a flow rate of $12.7 \text{ m}^3\text{s}^{-1}$</td>
</tr>
<tr>
<td>Consistency</td>
<td>A measure of the conflicts with redundant data</td>
<td>A pressure of 1200 bar in the ABC system is also stored 1200 bar in the XYZ system.</td>
</tr>
<tr>
<td>Entirety</td>
<td>A measure of the quantities of entities created</td>
<td>All well interventions made were recorded and stored.</td>
</tr>
<tr>
<td>Breadth</td>
<td>A measure of the amount of information captured about an object or event</td>
<td>All information about a specific injection regime is captured including duration, start and stop time, injected material, etc.</td>
</tr>
<tr>
<td>Completeness</td>
<td>A measure of information captured within a specific entity</td>
<td>UWI, status, field, well name are all stored for each well</td>
</tr>
<tr>
<td>Uniqueness</td>
<td>A measure of unnecessary information replication</td>
<td>Well Header is stored once for each well</td>
</tr>
<tr>
<td>Interpretability</td>
<td>A measure of semantic standards being applied</td>
<td>A date is stored as 11 June 2002</td>
</tr>
<tr>
<td>Freshness</td>
<td>A measure of how current a record is</td>
<td>Well status represents current status</td>
</tr>
<tr>
<td>Precision</td>
<td>A measure of exactness</td>
<td>The timestamp for Historian data is precise (e.g. 13:42:00.015627 )</td>
</tr>
<tr>
<td>Depth</td>
<td>A measure of the amount of history that is retained</td>
<td>A complete history of interventions, modifications, shut-ins, simulations, well tests is stored for each well</td>
</tr>
<tr>
<td>Integrity</td>
<td>A measure of validity relative to another item</td>
<td>Measured depth should be greater or equal to total vertical depth</td>
</tr>
</tbody>
</table>
Quality drives business application

Size of data set compensates for quality  More data just means more noise

**Strategic**  
100 – 1000 m  
- 100s of wells  
- Discontinuous seismic surveys  
- Monthly production

**Tactical**  
10 – 100 m  
- 10s of wells  
- 100s of sidetracks  
- Reservoir models

**Operational**  
1 – 10 m  
- 10s of wells and sidetracks  
- 1000s of petrophysical measurements

**Basin** (portfolio mgmt)  
Basic CRS and projection checks. Units? Duplicates? Nomenclature?

**Reservoir** (well planning)  
Reduced standard vocabulary, numerical validation at file level

**Formation** (geo-steering)  
Knowledge of tool string, muds, borehole reports, petrophysical database
Let’s take a close look at your well logs…
Can we fix it? (Yes we can)

- Better understanding of the problem and hence the business requirements
- Stronger Master Data Management – ensure data is indexed to the granular level
- Better Metadata Management – build cross-domain catalogues
- Quality and uncertainty metrics can be at the measurement level – storage is cheap

Data should be corrected for:
- Pull rate
- Vertical location
- Angle
- Numerical range
- Missing data
Compensating for missing data at the measurement level

\[ V_p = f(\text{depth, ILD}) \]

\[ V_s = f(V_p, V_{sh}) \]

\[ \rho = f(V_p, V_{sh}) \]

<table>
<thead>
<tr>
<th>Well</th>
<th>( V_p )</th>
<th>( V_s )</th>
<th>( \rho )</th>
<th>( R_t )</th>
<th>( V_{sh} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>W1</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>W2</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>W3</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

LAS Tops

LAS
Rock Property Zone Averages
Does the ML work?

Yes!
Basic user interface with standard BI tool
Putting ML and Quality back into the Digitalisation context

You want to respond to operational challenges and strategic opportunities with the best decision possible.

You need to be able to contextualise new data in appropriate time frames.

Machine Learning (and AI) is the way in which this is done at scale.

Your data quality needs to be awesome, or at least fit for purpose.

You need to change your culture, organisation, process and tools.

You need to stop storing your data in silos, and leaving provenance and quality to someone else.
Delivering on a digitalisation vision

**Organisation:** Define and establish an Upstream analytics capability to orchestrate and prioritise business-aligned data management. This should evolve into a Chief Data Office.

**People:** Train people in Agile methodologies to provide domain data leadership. Educate data domain and IT teams on how data management works in industries further down the digitalisation road. Align on concepts and vocabulary.

**Processes:** Data ingest and contextualisation needs close alignment with business requirements. Ensure strong governance of data management and utilisation capabilities.

**Quality:** Perform an audit of data quality and implement required data quality procedures to ensure a consistent framework of quality measurement, publication and continuous improvement. This will drive trust in data across an organisation and deliver on a digital vision.
Any questions?

You could buy a copy of this awesome book...

ENHANCE OIL & GAS EXPLORATION
WITH DATA-DRIVEN GEOPHYSICAL AND PETROPHYSICAL MODELS

KEITH R. HOLDAWAY
DUNCAN H. B. IRVING