TERADATA

The Rules of Subsurface Analytics

Jane McConnell, Practice Partner Oil and Gas, Teradata

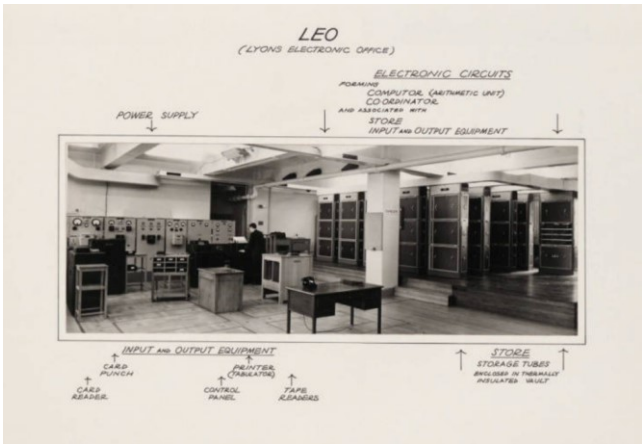
DEJ KL, 4 October 2017

Agenda

- Why subsurface analytics is different
- The Rules
  - Rule 1: Right People
  - Rule 2: Right Platform
  - Rule 3: "Good Enough" Data Management
  - Rule 4: Agile Approach
  - Rule 5: Business Buy-in
- Recap

Why Subsurface Analytics is different

1940's business computing: where it all started for business analytics



TERADATA

But for us, the IT/OT divide happened

TERADATA.

Subsurface Data Management Culture

Custodianship
v
Stewardship

Custodian

- Controls access
- Avoids unknown data
- Transfer data
- Avoids risk
- Hates change
- Acquires knowledge
- Creates walls

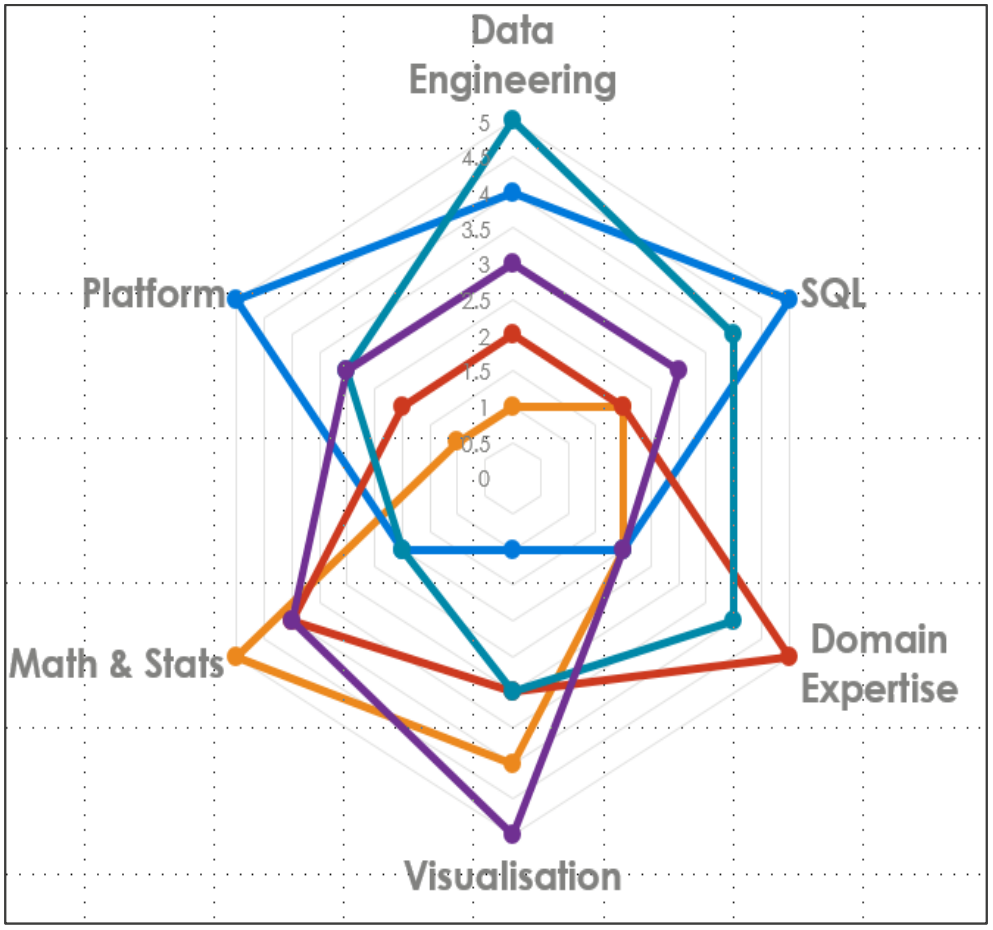
Steward

- Shares data
- Celebrates variety
- Enables access
- Embraces risk
- Owns change
- Shares insights
- Teaches governance

Rule 1: Right People

Right People, plural. And T-shaped.

- Too many disciplines for any one person to know it all
- "T-shaped people" who go wide across many disciplines but deep into their specific domain
- Need outstanding data management and data engineering skills (and culture)
- Need platform expertise for sustainability and deployment
- Need Subject Matter Expertise



Analytics / data science workflow

9

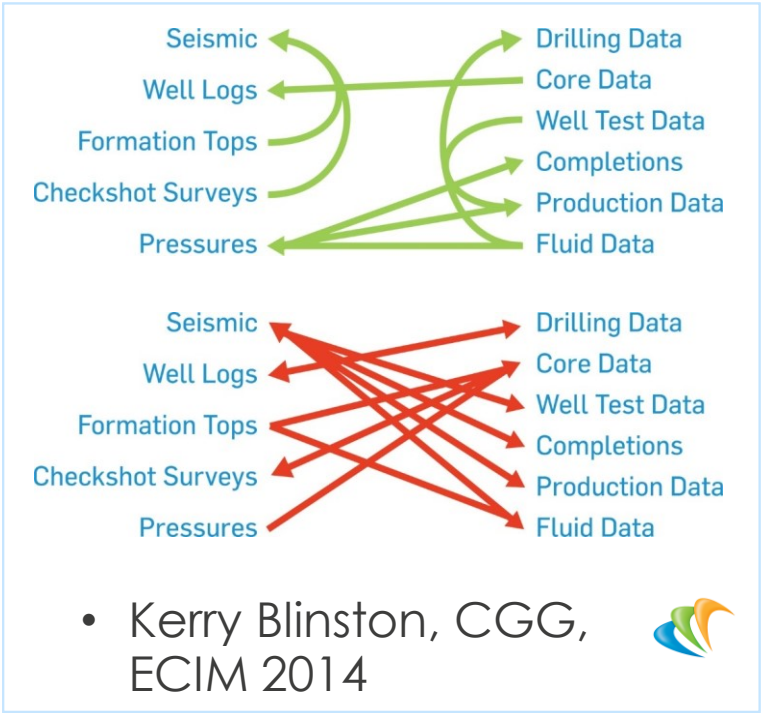
Here's one we did earlier in Malaysia

- Working closely with the customer
  - Subject matter expertise
  - Source system expertise
- Teradata Subject Matter Expert
- Data management skills
- Data platform skills
- Coding skills
- Data science skills
- Frontend/visualisation skills

Rule 2: Right Platform

The problems with existing data stores

- "Knowledge development" applications come with import filters for specific file types and specific tasks
- Data is modelled logically for well-defined (and hence brittle) processes that may not reflect all (or even any!) use cases
- Only "perfect" data can be imported into applications or schemas

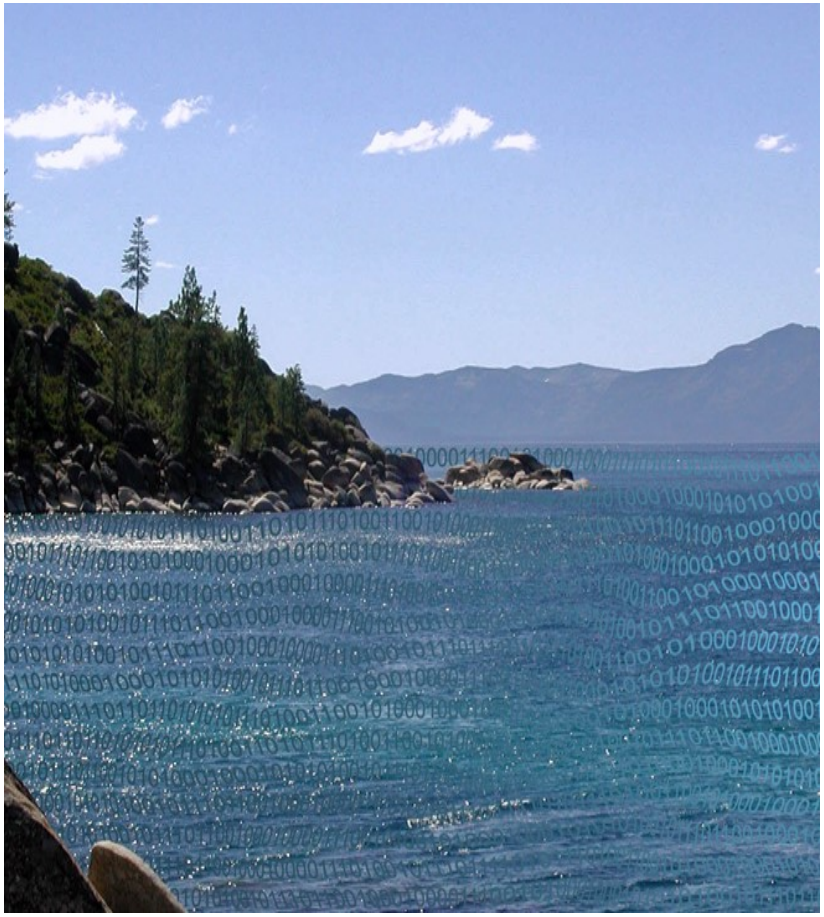


- Kerry Blinston, CGG, ECIM 2014

New data types, or new combinations, challenge all of this

Build a new platform that all disciplines can use

- If we don't give them a purpose-built platform for analytics, we will be in Desktop/Excel Hell.
- A platform that
  - Accepts data from any discipline
  - Makes it easy for data scientists to use their tools – R, Python etc
  - Provides the right level of governance and data quality
- Call it a Data Lake, or call it an Analytical Data Platform
  - Just build something where they can get data and ideally write new data back



Rule 3: Good enough Data Management

“Good Enough” Data Management

- Stewards mentor and support “citizen data management”
- Everyone cares about the data and its quality
- Everyone can do something about it when they find bad data
- Data governance is a function of data value

“Good Enough” means:

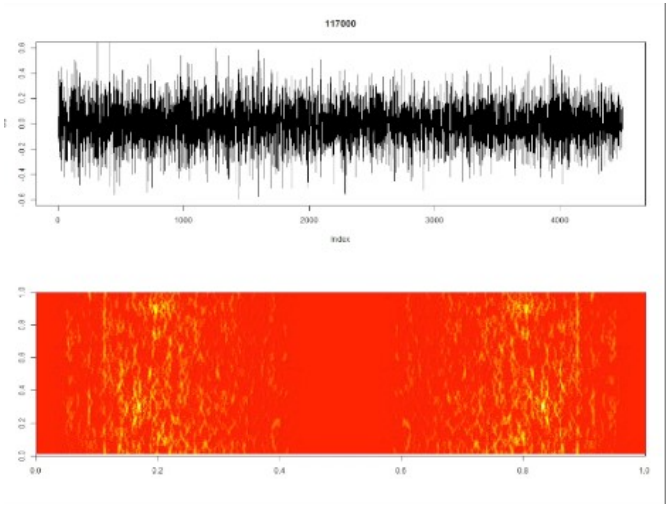
- **Good:** don’t compromise on quality
- **Enough:** don’t boil the ocean

What should data look like?

Don't be a data hoarder!

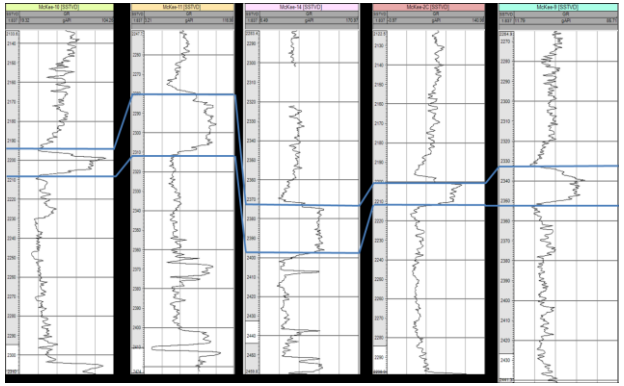
Why not store data at a granularity good enough to extract value?

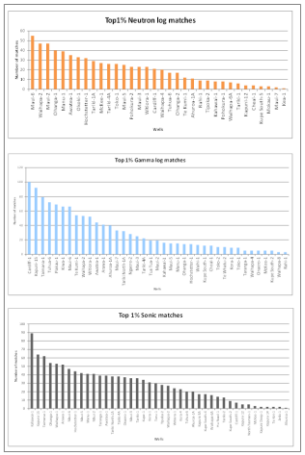
- Granular enough
- Dimensioned (time, space) enough
- Resample, interpolate, aggregate
- Profile data so you can get an idea of it quickly



“Difficult file formats” (Multi-structured data)

- Parse out the measurement data
- Link it through time and space
- Relate using metadata and master data
- Form a view on whether a hypothesis is worth developing





GOOD ENOUGH!

Dealing with unstructured data

Text

- Language
- Typos
- Consistency
- Quality

Use simple characterisation tools to understand what is in the data

Don't try to build a whole text input and cleansing framework

3203	recalibration	13
3204	receiver	8
6895	receiving	9
1273	recheck	7
6896	rechecked	9
6897	rechecks	8
6898	recleaning	10
3192	re-cleaning	11
6899	recomissioned	13
3206	recomissioning	14
3207	recommended	11
6900	recommission	12
3208	recommissioned	14
6901	recommissioning	15
6902	recorded	8
6903	recover	7
3210	recovery	8
227	rectification	13
3212	rectified	9
3213	rectify	7
6906	rectifying	10
3216	redivert	8
6907	reduce	6
3217	reduced	7
6910	reducer	7
3218	reducing	8
3219	reduction	9
3220	reenergise	10
3221	reenergised	11
3222	reestablished	13

TERADATA

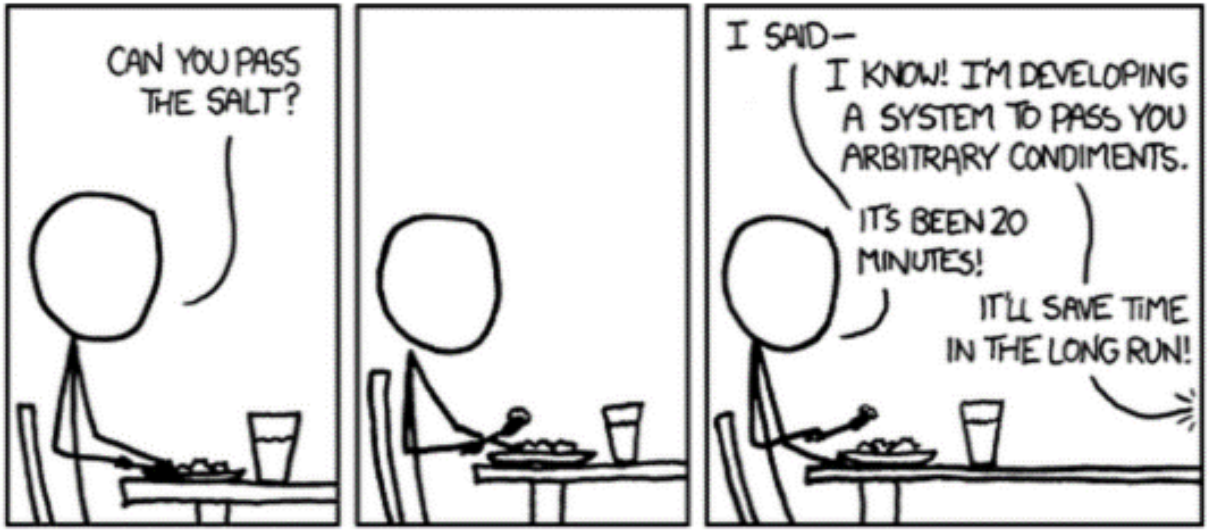
Rule 4: Agile Approach

Agile, Scrum, DevOps, Interactive Visualisation

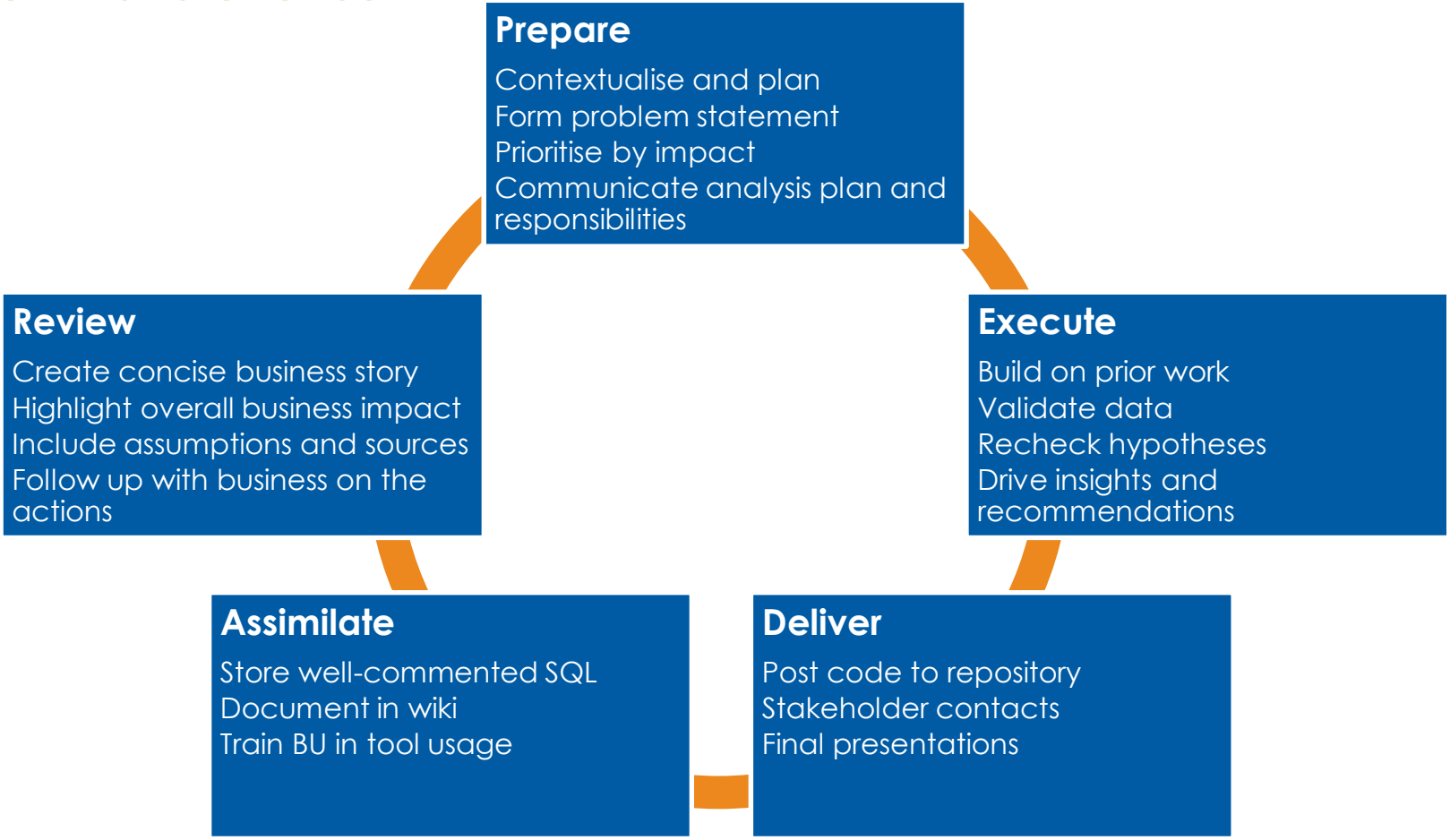
What not to do

Common resource-sinks:

- Point solutions
- Technology projects
- Waterfalls
- Brittle data modelling
- "We want ML/AI!"



Iterate.
One project at a time.
Deliver value often.



Agility needs the right mindset

Working together

Willing to take risks

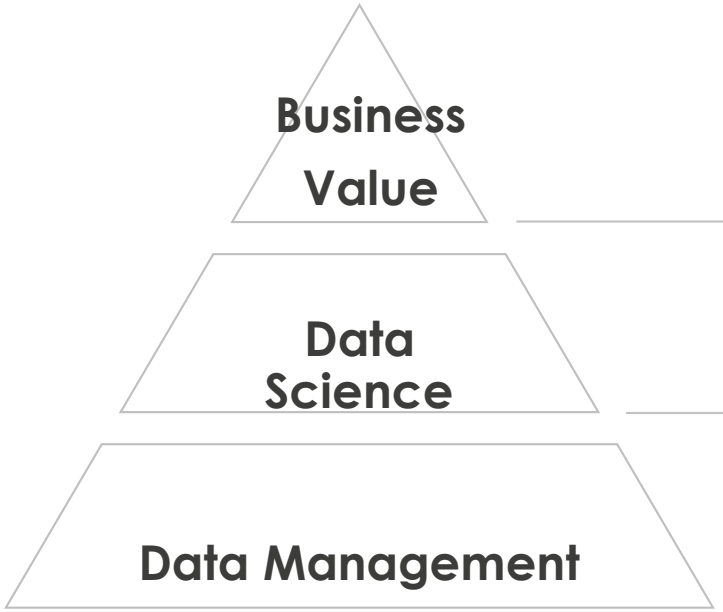
Proactive

Speed / Time-to-insight

TERADATA.

Rule 5: Business Buy-in

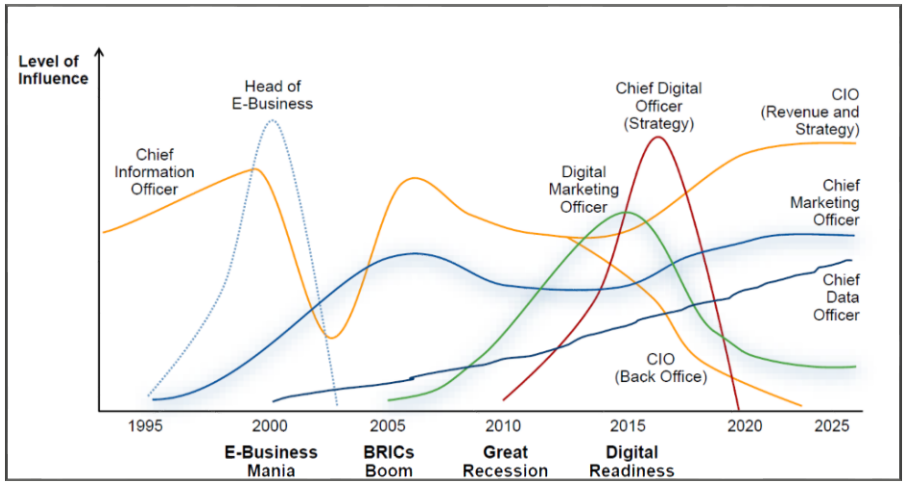
Data management to enable business-aligned data science



- What financial, operational or environmental impact are you delivering?
- What techniques, functions, workflows and skills are required?
- What data is required and in what form?

Business-focused data management

- Embed data ownership in the business units
- Engage with business leadership to plan, budget and deliver data-driven initiatives
- Define and drive data exploitation strategy
- Understand data value and leverage high value data for business impact



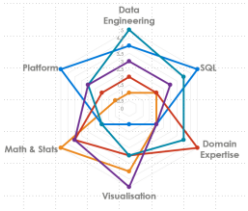
Source: Andrew White, Gartner 2017

…do we need a
Chief Data Officer?

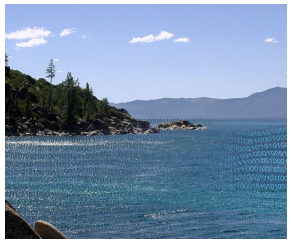
TERADATA

Recap

The 5 Rules – all equally important

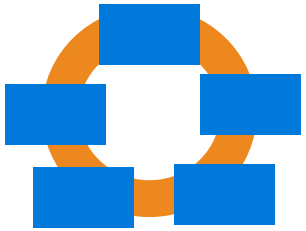


Right People

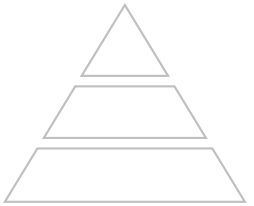


Right Platform

Good Enough Data Management



Agile Approach

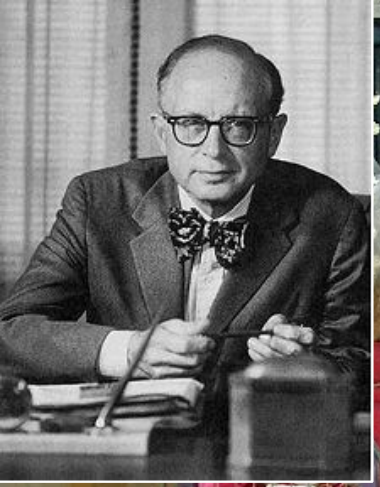


Business Buy-in

TERADATA.

Stated another way - Data Management 2.0

Stop doing	Keep doing	Start doing
• Brittle data management • Silos • Disposable data science • Transfer and analysis in Excel	• Applying domain expertise • High levels of governance • Driving data quality • Learning	• Aligning with business • Applying context • Data profiling • Enriching data • Applying critical thinking



Closing thoughts

"Trying to plan for the future without a sense of history is like trying to plant cut flowers."

Daniel J Boorstin (1914 – 2004)
12th Librarian of the US Congress 1975-1987

Jane McConnell
Practice Partner O&G , Industrial IoT Group
Thing Big Analytics, a Teradata Company
Jane.mcconnell@thinkbiganalytics.com
+47 98282110

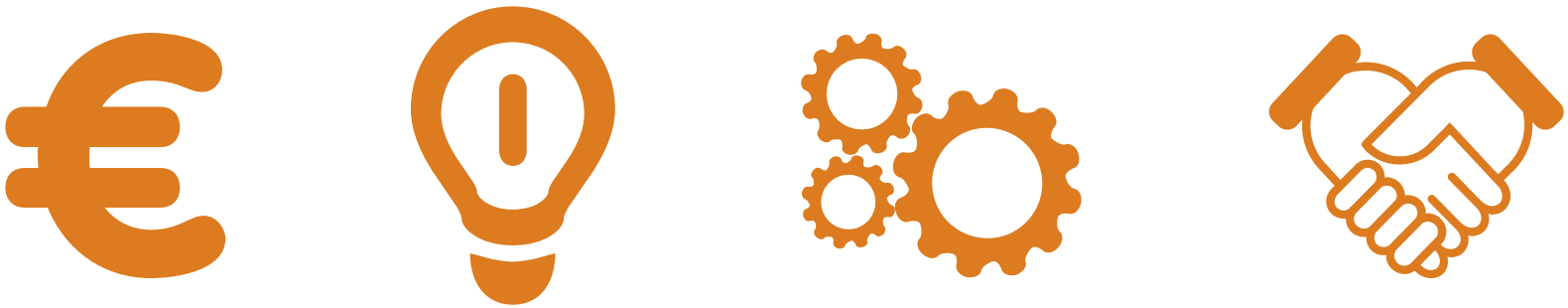
My blog on Forbes
My blog on Teradata.com
Follow me on Twitter @jane_mcconnell
My profile

TERADATA

Jane McConnell
Practice Partner O&G , Industrial IoT Group
Thing Big Analytics, a Teradata Company
Jane.mcconnell@thinkbiganalytics.com
+47 98282110

My blog on Forbes
My blog on Teradata.com
Follow me on Twitter @jane_mcconnell
My profile

What is the point of data science?



Data science, also known as data-driven science, is an interdisciplinary field about scientific methods, processes, and systems to extract knowledge or insights from data in various forms, either structured or unstructured, similar to data mining.

Wikipedia, 2017

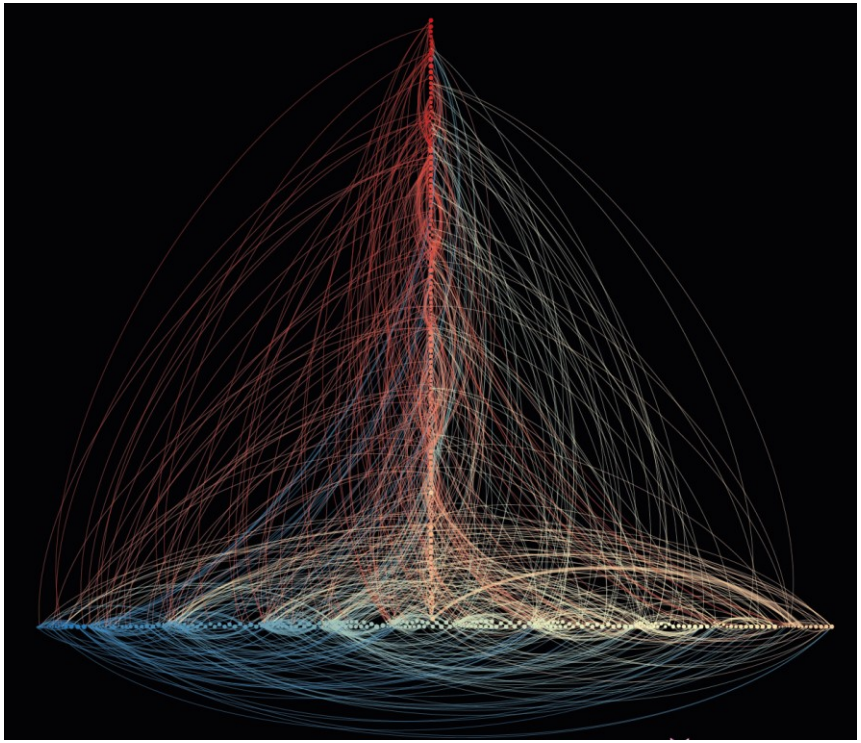
Data science workflow

3
3

TERADATA.

Not just machine learning

- Finding relationships with complex data sets
- Characterising behaviour and understanding the **demographics** of data
- It can be applied to:
  - Data profiling and QC
  - Data preparation
  - Data mining
  - Operational processes
  - Data art



How other industries grew an analytics culture…