

Defining BDRA

Introduction to the BDRA symposium “Foundations of BDRA”



Coen van Gulijk

First....

Congratulations!!!

University of
HUDDERSFIELD
Institute of Railway Research

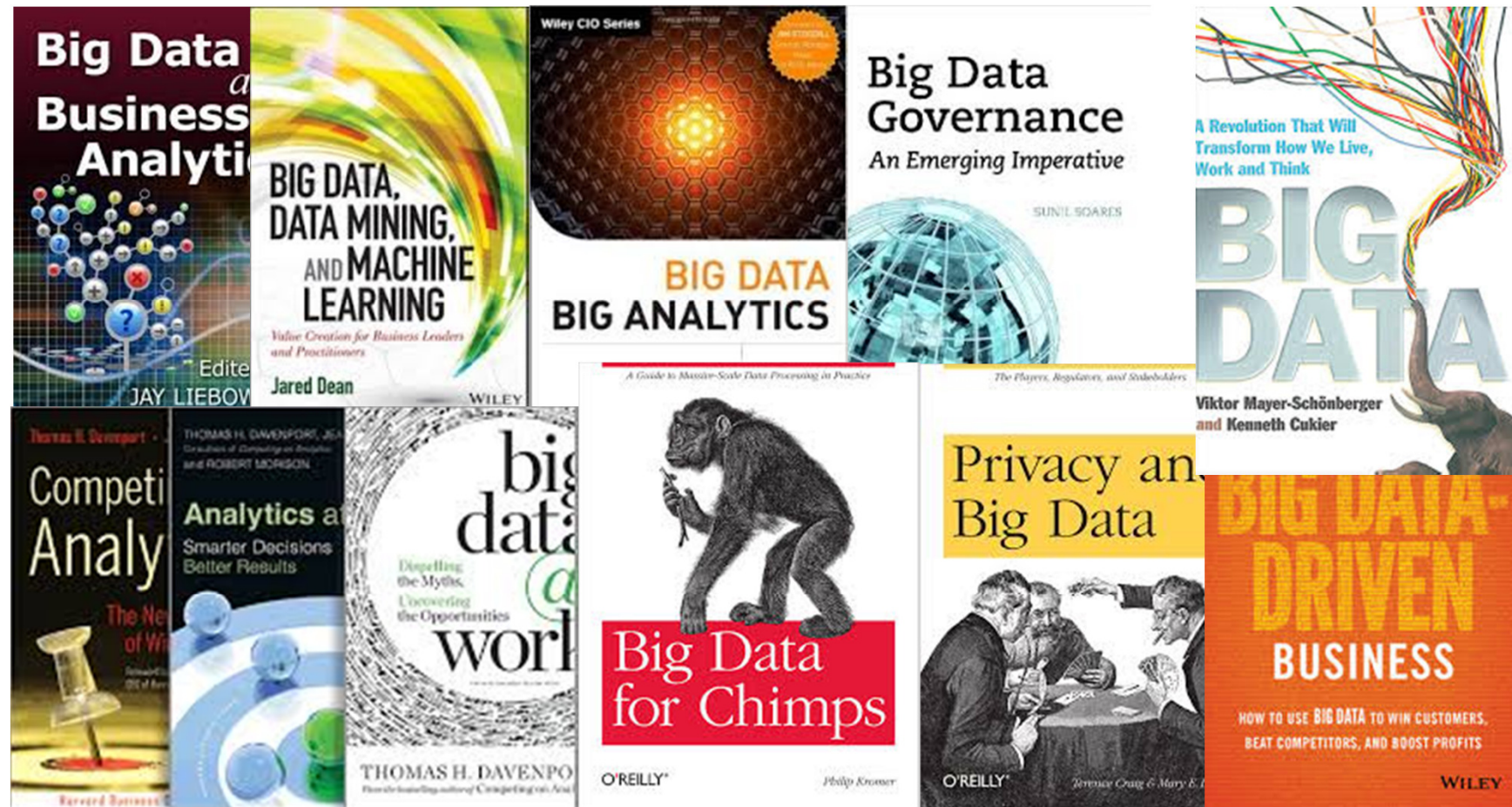
Today is 28 April

The annual World Day for Safety and Health at Work



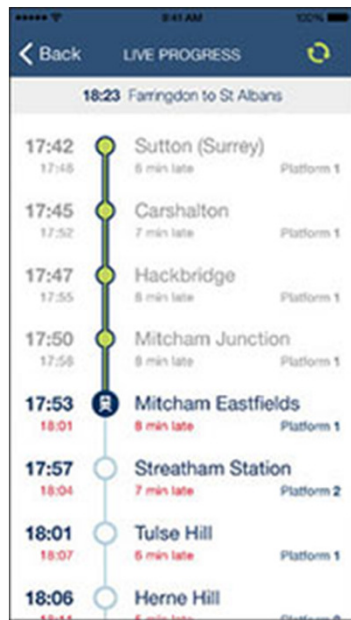
Are you a believer?

University of
HUDDERSFIELD
Institute of Railway Research



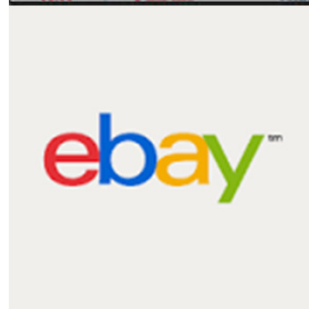
Are you a user?

University of
HUDDERSFIELD
Institute of Railway Research



18:23 Farringdon to St Albans

17:42	Sutton (Surrey)	Platform 1
17:45	6 min late	
17:45	Carshalton	Platform 1
17:52	7 min late	
17:47	Hackbridge	Platform 1
17:55	8 min late	
17:50	Mitcham Junction	Platform 1
17:58	8 min late	
17:53	Mitcham Eastfields	Platform 1
18:01	8 min late	
17:57	Streatham Station	Platform 2
18:04	7 min late	
18:01	Tulse Hill	Platform 1
18:07	6 min late	
18:06	Herne Hill	



You are in...

University of
HUDDERSFIELD
Institute of Railway Research

Whether you like it or not...

Film as an analogy for Big Data

University of
HUDDERSFIELD
Institute of Railway Research

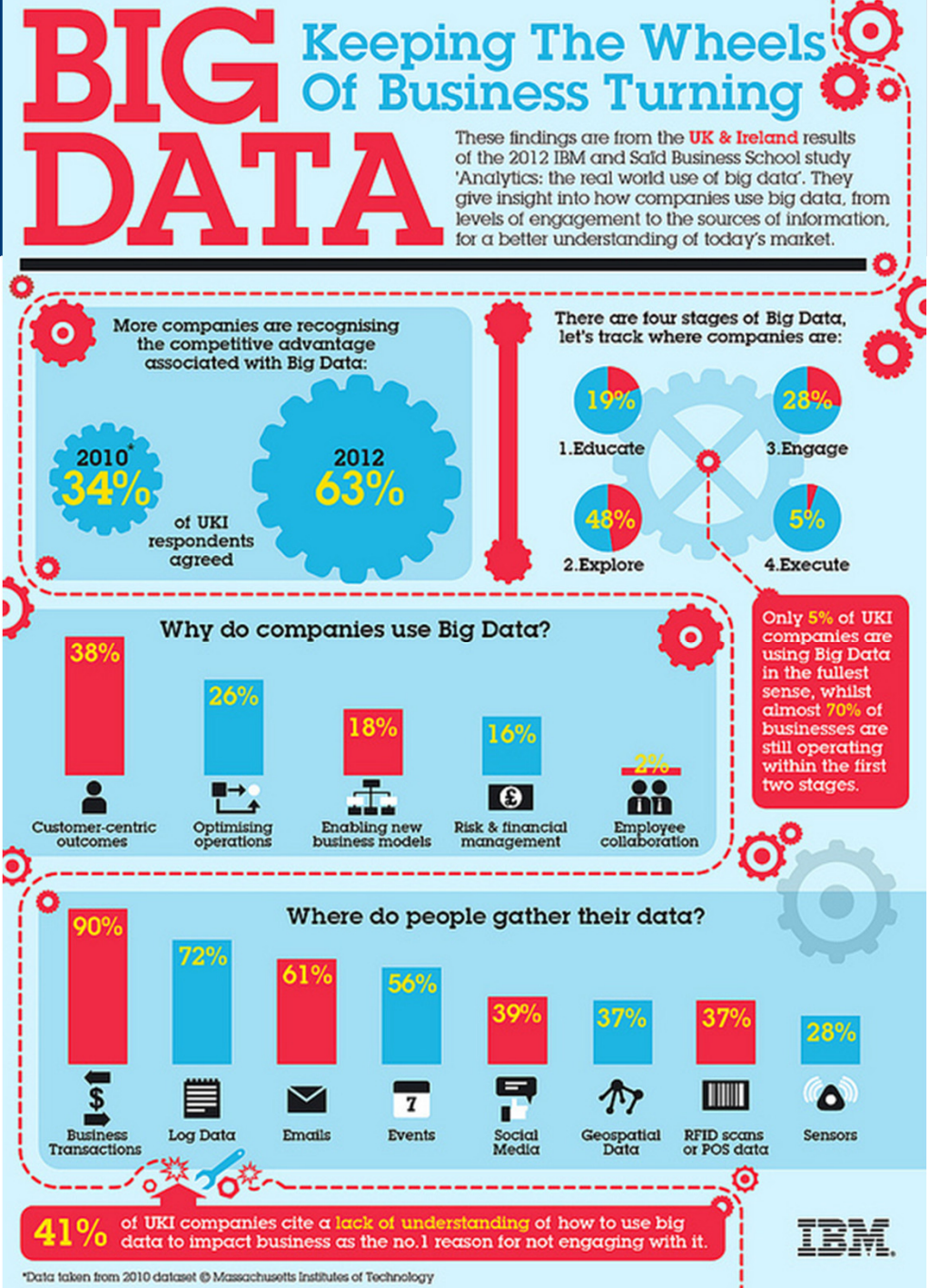


Anecdotal success stories

- *Amazon* adds 30% revenue by second guessing your next buy
- *Google* translate uses all translations known on internet to enable their translator to have 94% accuracy (?)
- *IBM* Watson helps oncologists make better informed decisions
- *NSA* has detected 152 terrorist plots
- *Target* knows when you are pregnant
- *UPS* USA saved millions by using predictive analytics on their fleet of 60,000 vehicles

Added value BD

- Customer data
- Optimizing operations
- New business models
- Risk & financial mgmt.
- Employee collaboration



What about GB railways?

I would have wanted to talk about the Digital Railway. [...]
I'm a passionate believer in the opportunity we have to lay out the blueprint for a technological transformation of rail in this country; a railway where in just 15 years or so we could lead the world in digital train control, delivering more capacity, reliability, speed and safety all at lower cost and with a smaller environmental footprint.

Mark Carne Chief executive, Network Rail
George Bradshaw Address 2015 Network Rail

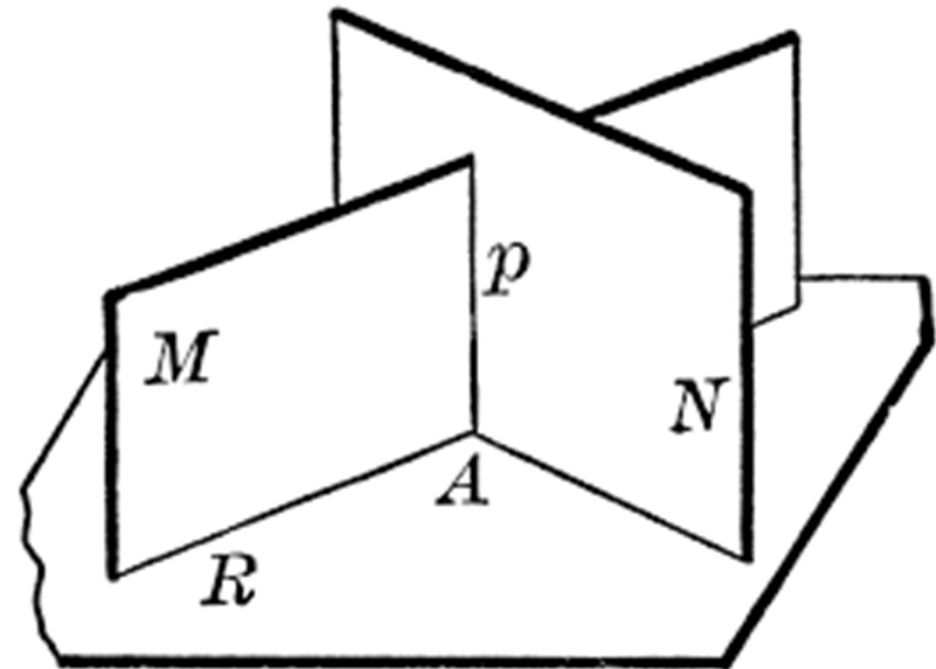
Where does that leave us?

BDRA for GB railways intersect at A where:

- Plane M is data analytics
- Plane N is safety and risk
- Plane R is digital railway



University of
HUDDERSFIELD
Inspiring tomorrow's professionals

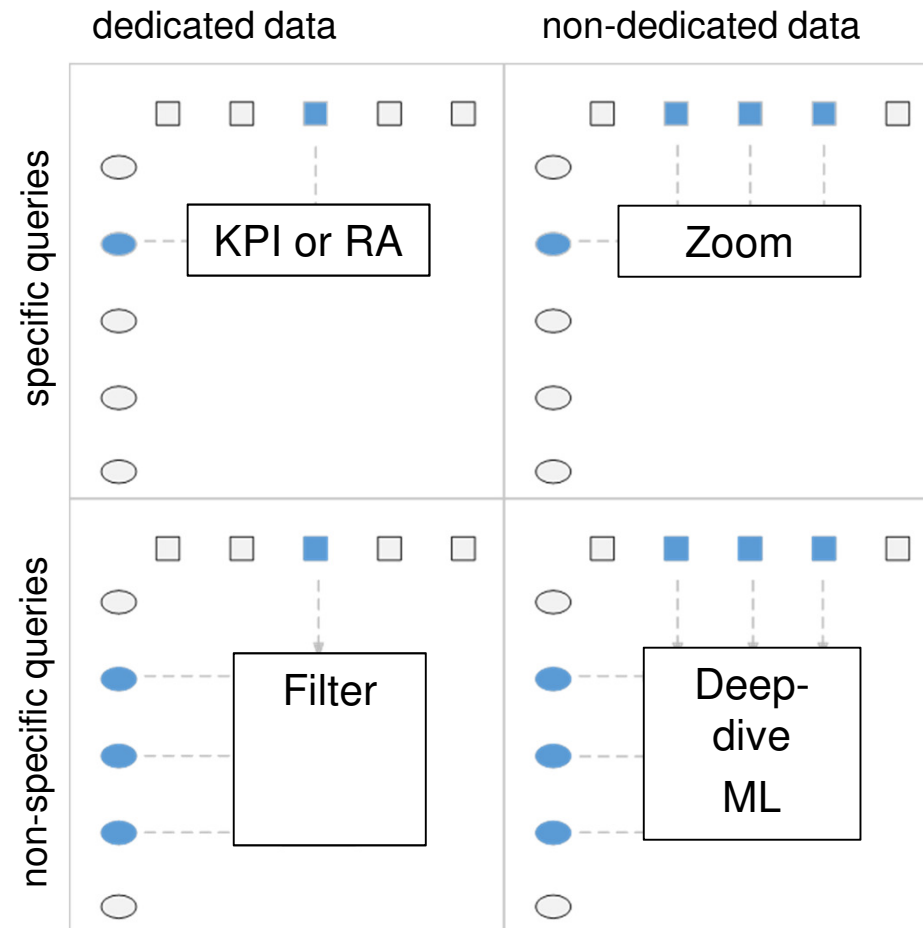


Who is out there on this intersection?

- We teamed up with RSSB
- NR is working on solutions
- TOCs are introducing data-centered solutions
- We found Lumeå University
- TU Delft expressed interest
- There are pioneering software companies in UK
- ...

Can we classify the efforts?

- Types of data
 - Dedicated: or purpose built (SMIS)
 - Non-dedicated: whatever you can grab
- Types of queries
 - Specific: purpose built (e.g. KPI/Risk An.)
 - Non-specific: inferring correlations with minimal guidance



Based on: Shneiderman, B. (1996). The Eyes have it: a task by data type taxonomy for information visualizations. In Proceedings of the 1996 IEEE Symposium on Visual Languages (pp. 336–343). Univ of Maryland, College Park, United States: IEEE.

So far on the Positioning of the BDRA program

- Test the promise of BD
- Joint effort HUD RSSB
- Intersect Big Data / Digital Rail / Safety & Risk

About the BD and safety

When Captain Picard talked to Data...
he actually got sensible information
(well, used to, a long time ago in a galaxy far away...)

For the rest of us...
it's just hard work



We can easily be carried away



Do not board this train;
it has a 13% chance of
derailing in York

The truth is...

Despite loads of visions in Big-Data books

- There are no off-the-shelf products that can do BDRA
- Developing successful machine learning applications requires a substantial amount of “black art” that is difficult to find in textbooks¹
- Access to relevant data is not guaranteed

That means we have to delve into black art.

But armed with the safety and risk toolbox.

Let's ponder on that for a while.

1: Domingos P (2012) a few useful things about machine learning. *Comm. Of the ACM* **55** (10): 78 – 87.

Defining BDRA (1)

3V

- Volume - lots of data (N=all)
 - Velocity - quickly stacking up
 - Variety - messy, can be all kinds of data
-
- Messy and quick is scary; risk analysts don't like that
 - N=all is an interesting notion
 - Julian Stow: RAATS project

Defining BDRA (2)

- Data analytics - mathematical software tools to infer decisions/knowledge from data
- This is not so far away from today's risk analysis
- In this light BDRA is not so new
- Paul Swuste's history overview

Defining BDRA (3)

- Degradation of causality - correlation is good enough
- Safety and risk analysts hate this
- How can you convince anyone to invest in safety if you are not sure that you are grabbing the problem by the root?
- Ben Ale's presentation

Defining BDRA (4)

- Data - the stuff that floats around the internet
- But unlocking it for safety requires machine learning
- Peter Hughes: Close Call project

Defining BDRA (5)

- Information - somehow useful for safety & risk
- But that requires massive computer power
- Violeta Holmes: High-Performance Computing

Defining BDRA (6)

- Railway safety
- RSSB's prerogative
- Marcus Dacre: changing SMIS database in more flexible big-data-like structures

So what is it?

BDRA is a software system that:

- extracts information from mixed data sources;
- processes them quickly to infer and present relevant safety management information;
- combines applications to collectively provide sensible interpretation;
- uses online interfaces to connect the right people at the right time

In order to:

- provide decision support for safety and risk management.

And we do not start from Ø

But elements of it are investigated around the UK

- Analyzing the Close Call database
- RAATS (analyzing NR data-feed)
- SMIS+
- RFID tagging of rolling stock
- OTMR recorders
- Analysis of Tweets
- GPS data on trains
- Big Data/Data Analytics as a research area
-

What's important?

- Software architecture
- Collating and combining tools into integrated system
- Visualization techniques
- Lexicon and Ontology
- The dark art of Machine learning
- Access to data-bases
- Hardware
- User requirements
- And somewhere to start: RSSB's Safety Risk Model

We have a lot to do

Lets get on with it...