University of HUDDERSFIELD
Institute of Railway Research

RSSB

Estimating the Approach Frequency to Red Signals – A Big Data Approach

Yunshi Zhao, Julian Stow, Institute of Railway Research
Chris Harrison, RSSB

Inspiring tomorrow's professionals

THE AWARDS AWARD WINNER UNIVERSITY OF THE YEAR

theguardian UNIVERSITY AWARDS Winner 2013

THE AWARDS 2012 WINNER Entrepreneurial University of the Year

THE QUEEN'S AWARDS FOR ENTERPRISE

Introduction

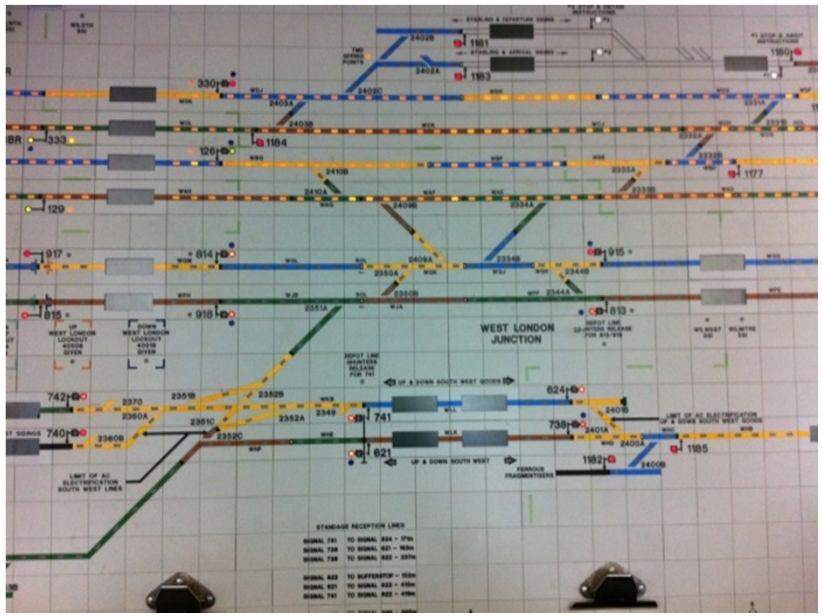
University of HUDDERSFIELD
Institute of Railway Research

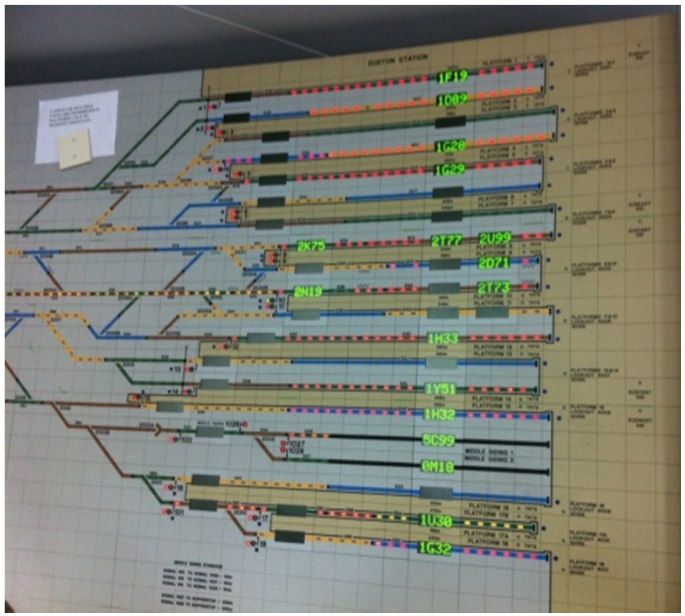
- The industry places great emphasis on understanding and controlling risks from SPADs
- Knowing the number of trains which approach signals displaying a red aspect is fundamental to:
  - Understanding SPAD risk at individual signals
  - Normalisation data for trending and benchmarking of SPADs
- Previous attempts have estimated the number of red aspect approaches nationally using driver surveys
  - These are considered to have significant shortcomings.
- Can industry data be used to improve this understanding……..?

Background

University of HUDDERSFIELD
Institute of Railway Research

- Knowing the location of trains, and the aspect displayed by signals is fundamental to the operation of the railway.
- A Train Describer (TD) on each signalling panel provides a description of each train in the berth it currently occupies

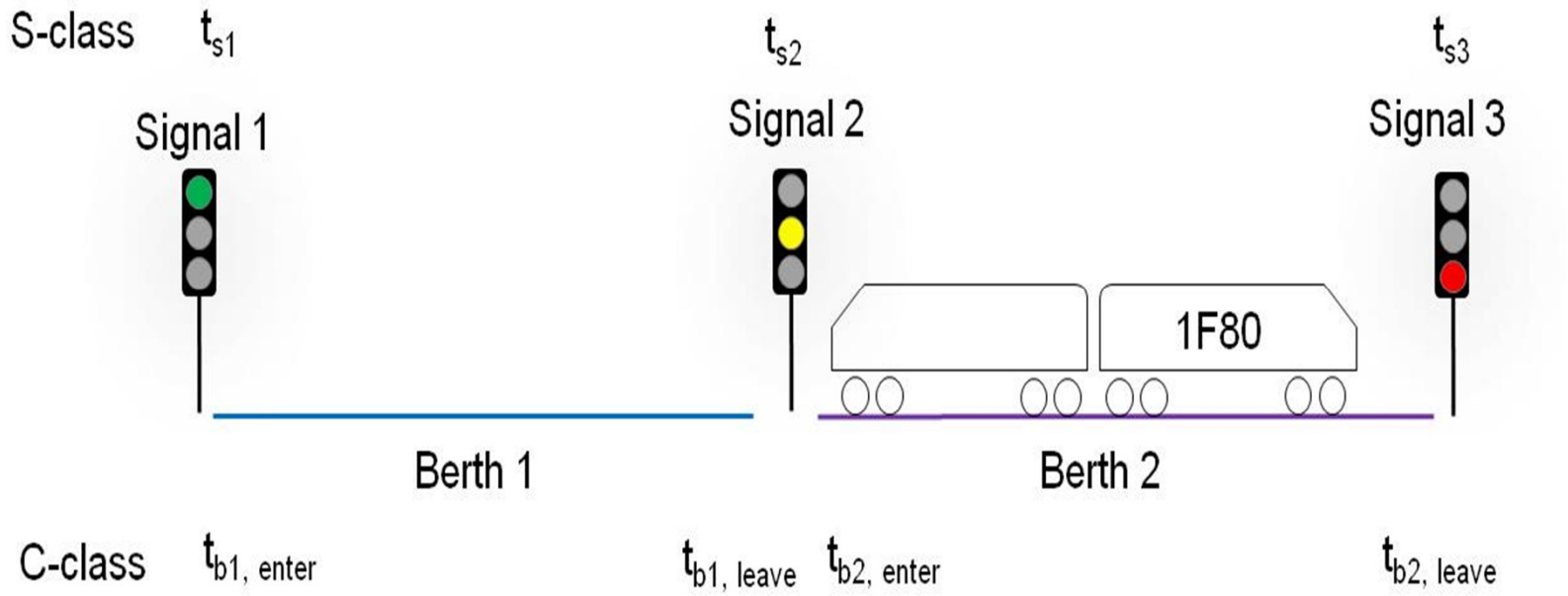




The Data

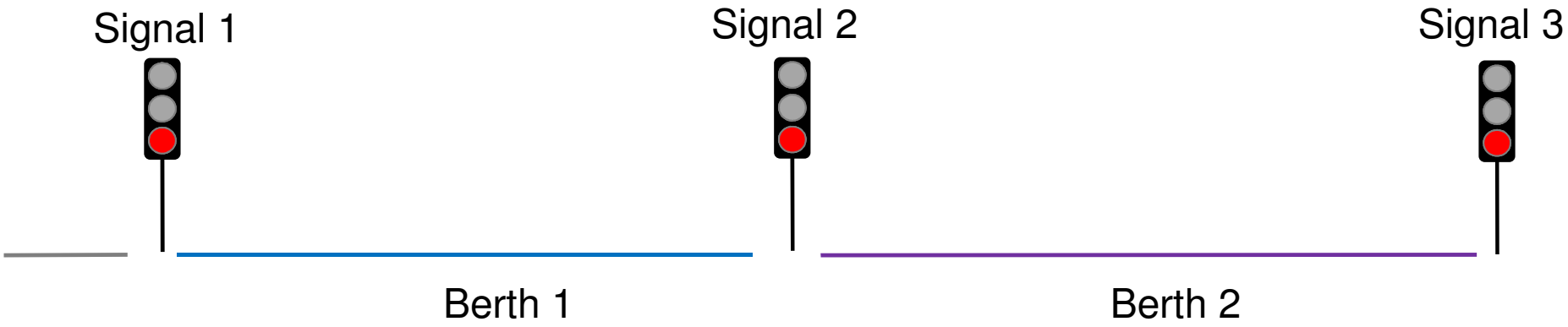
University of HUDDERSFIELD
Institute of Railway Research

- TD data is publicly available under Network Rail's open data initiative www.networkrail.co.uk/data-feeds
  - S-class messages give signal states (on/off)
  - C-class messages give train movement data through 'berths'



The Data

University of HUDDERSFIELD
Institute of Railway Research



S-class

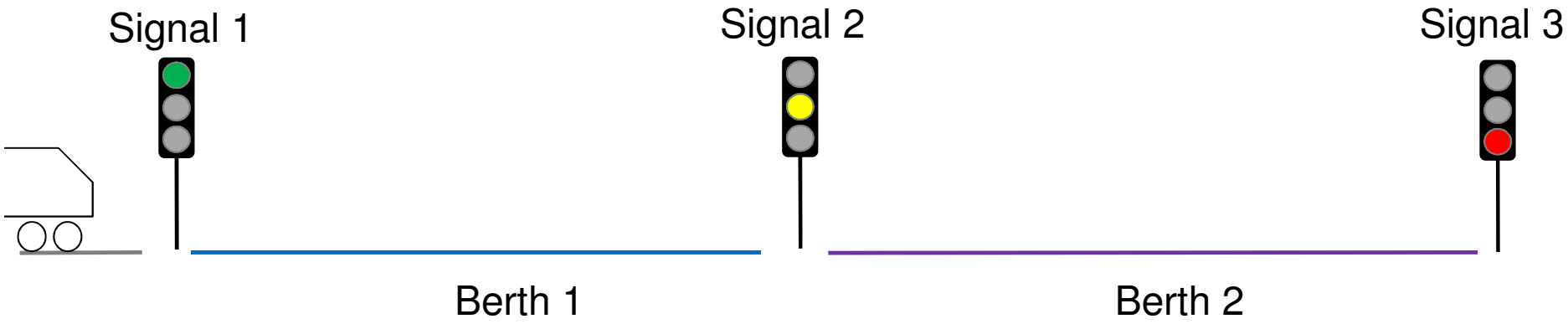
Time	Signal 1	Signal 2	Signal 3
16:45:00	1	1	1

C-class

	Berth 1		Berth 2	
Train	Enter	Exit	Enter	Exit

The Data

University of HUDDERSFIELD
Institute of Railway Research



S-class

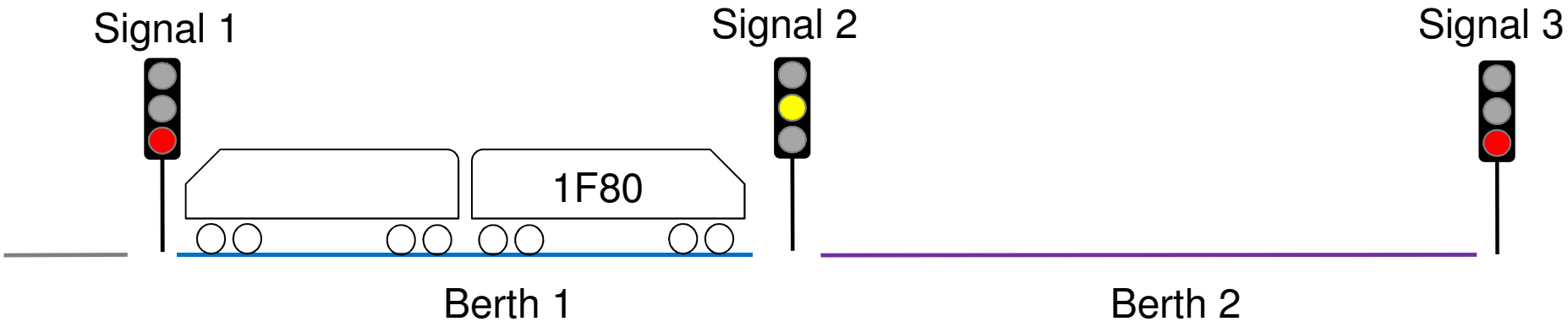
Time	Signal 1	Signal 2	Signal 3
16:45:00	1	1	1
16:46:00	0	0	1

C-class

	Berth 1		Berth 2	
Train	Enter	Exit	Enter	Exit

The Data

University of HUDDERSFIELD
Institute of Railway Research



S-class

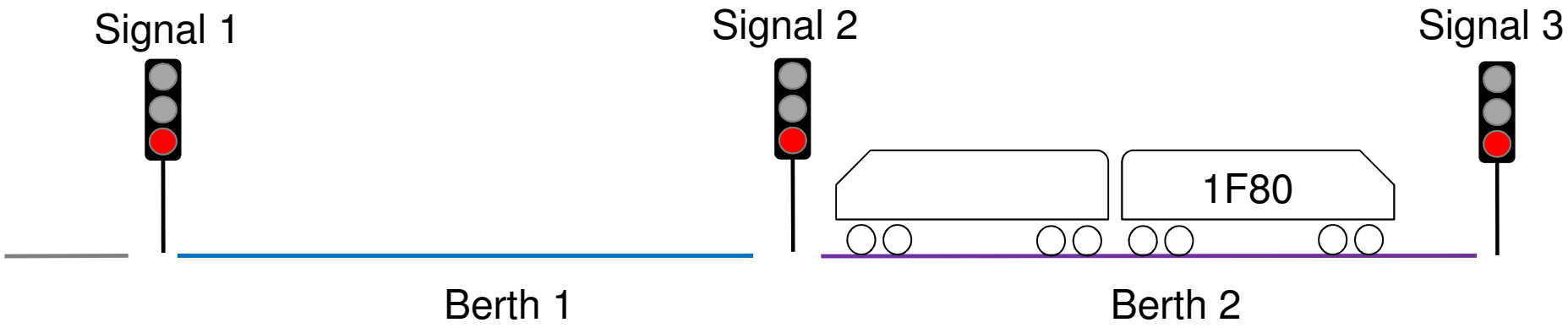
Time	Signal 1	Signal 2	Signal 3
16:45:00	1	1	1
16:46:00	0	0	1
16:50:00	1	0	1

C-class

	Berth 1		Berth 2	
Train	Enter	Exit	Enter	Exit
1F80	16:50:00			

The Data

University of HUDDERSFIELD
Institute of Railway Research



S-class

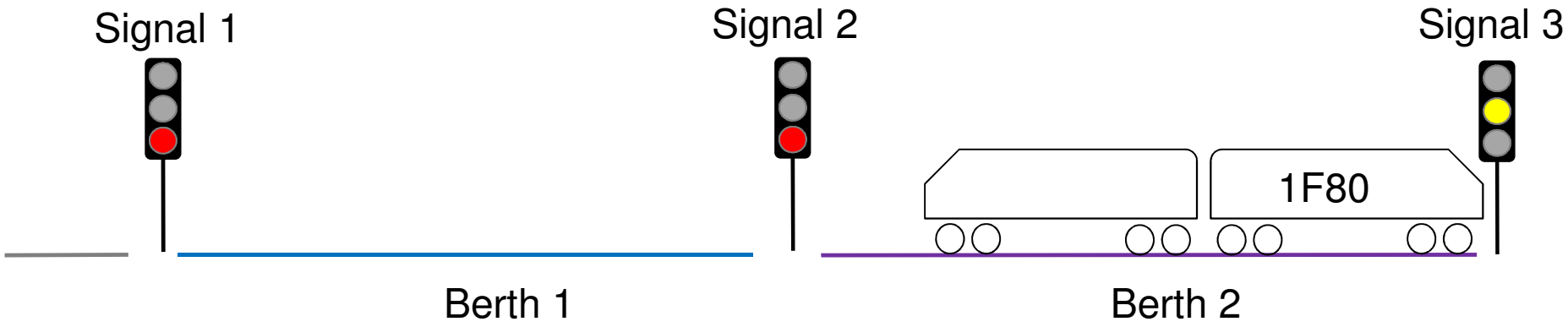
Time	Signal 1	Signal 2	Signal 3
16:45:00	1	1	1
16:46:00	0	0	1
16:50:00	1	0	1
16:53:00	1	1	1

C-class

	Berth 1		Berth 2	
Train	Enter	Exit	Enter	Exit
1F80	16:50:00			
1F80	16:50:00	16:53:00	16:53:00	

The Data

University of HUDDERSFIELD
Institute of Railway Research



S-class

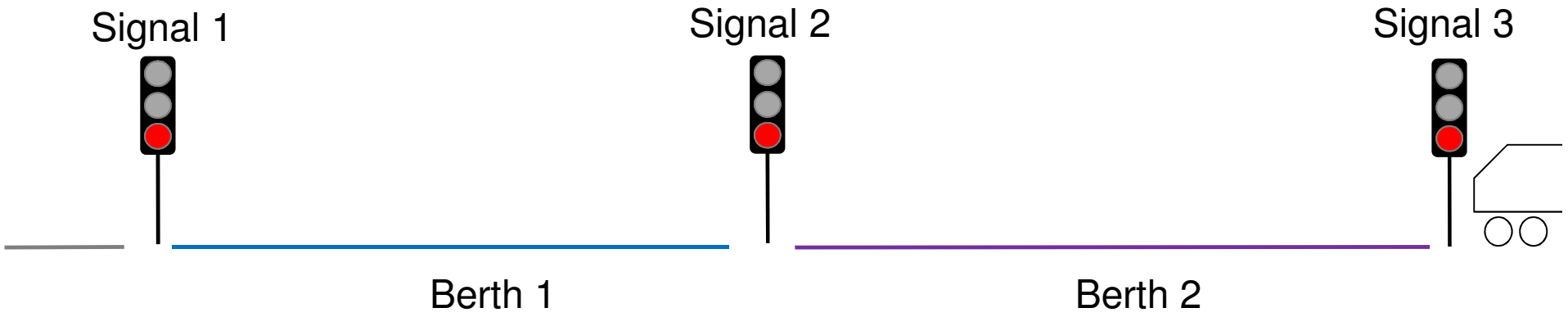
Time	Signal 1	Signal 2	Signal 3
16:45:00	1	1	1
16:46:00	0	0	1
16:50:00	1	0	1
16:53:00	1	1	1
16:54:00	1	1	0

C-class

	Berth 1		Berth 2	
Train	Enter	Exit	Enter	Exit
1F80	16:50:00			
1F80	16:50:00	16:53:00	16:53:00	

The Data

University of HUDDERSFIELD
Institute of Railway Research



S-class

Time	Signal 1	Signal 2	Signal 3
16:45:00	1	1	1
16:46:00	0	0	1
16:50:00	1	0	1
16:53:00	1	1	1
16:54:00	1	1	0
16:54:10	1	1	1

C-class

	Berth 1		Berth 2	
Train	Enter	Exit	Enter	Exit
1F80	16:50:00			
1F80	16:50:00	16:53:00	16:53:00	
1F80	16:50:00	16:53:00	16:53:00	16:54:10

The Data – Details and Caveats

University of HUDDERSFIELD
Institute of Railway Research

- TD coverage
  - 146 TDs provide train movement (C-class) messages
  - Only 61 TDs provide signal aspect (S-class) messages
- Signal IDs are not directly related to berth IDs
- TD S-class decoding
  - 7 out of 61 TD areas are included in the current decoding map
  - Not all signal elements are covered in these TD areas
- The data feed doesn't provide geographic locations, info about signal type etc.

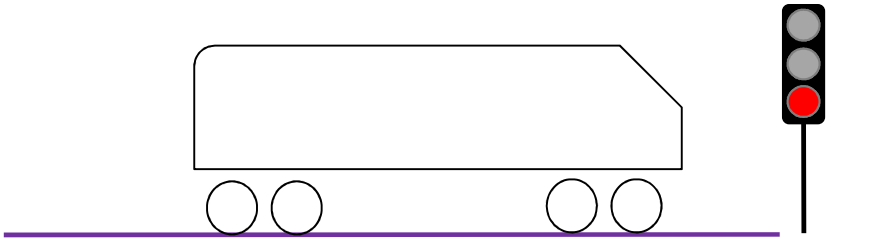
Is it Big Data?

University of HUDDERSFIELD
Institute of Railway Research

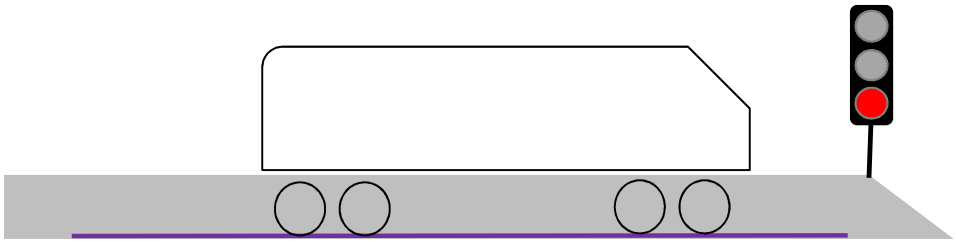
- Philosophically:
  - Old paradigm – small sample, limited data, extrapolate
  - New paradigm – get all the data, analyse, interpret
- Volume? ✓ Velocity? ✓ Variety? ✗
  - ~5.2 million signalling & train movement messages /day
- But:
  - It is still very structured
  - We don't really use the 'velocity'
  - It is probably only big-ish
  - Our analysis techniques are not really 'Big Data'

Analysis Model - Definitions

University of HUDDERSFIELD
Institute of Railway Research



CSS – Cleared Stopped (at) Signal



CBD – Cleared Before Departure

CAS – Cleared Approaching Signal
CNV – Cleared Not Visible

CBB – Cleared Before Berth

Data Flow

University of HUDDERSFIELD
Institute of Railway Research

Data Receiving & Storage

- Receive data from NR live feed
- Save the data in an SQL database

Data Processing

- Join Train Movement and Signal Status messages to provide information on the movement of every train and the aspect of the signal it approaches.

Analysis Model

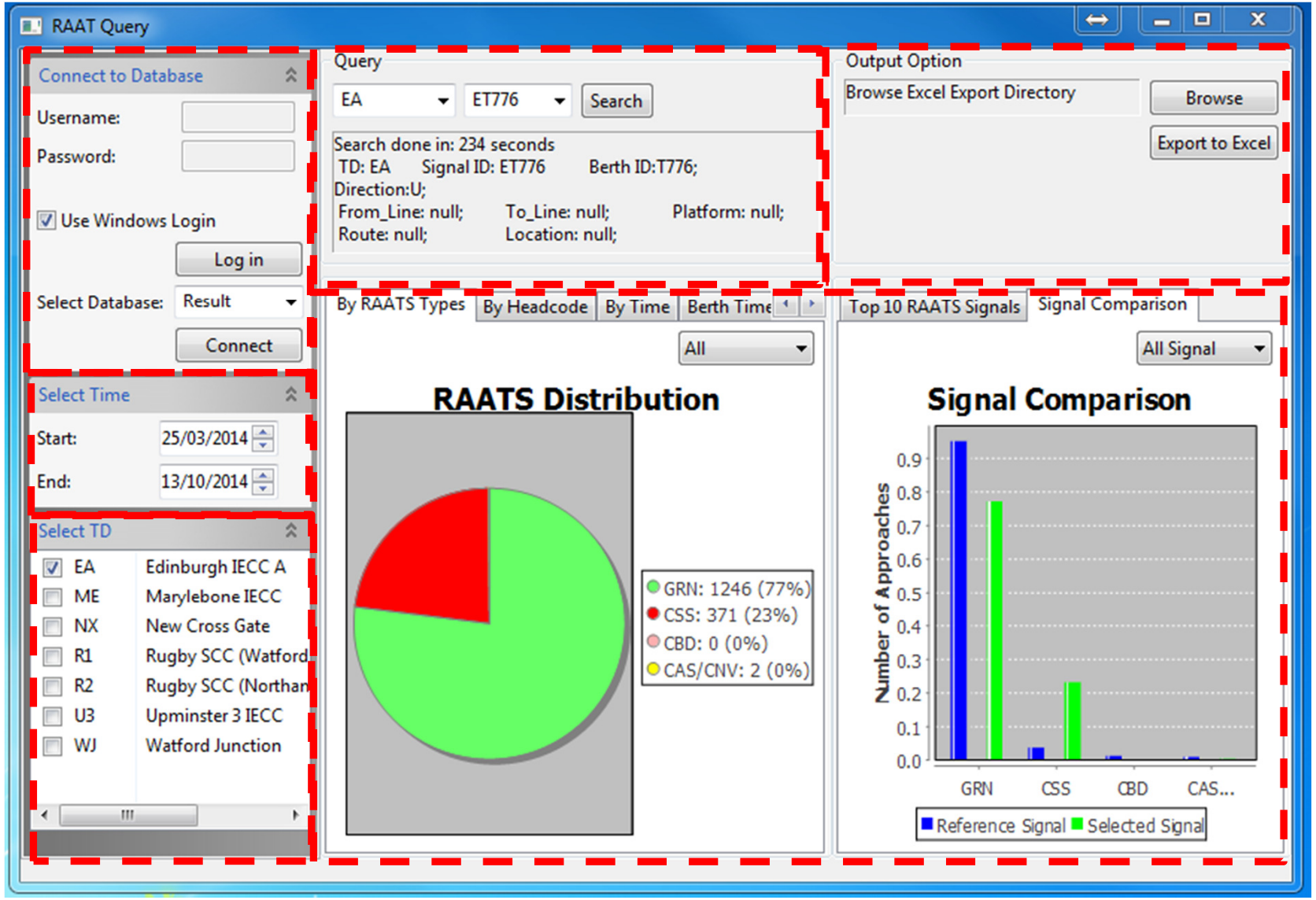
- Identify types of red signal approach incidents using the processed data.
- Collate data for individual signals, groups of signals etc.

GUI

- Present results to user in various formats
- Export data for external use or further analysis

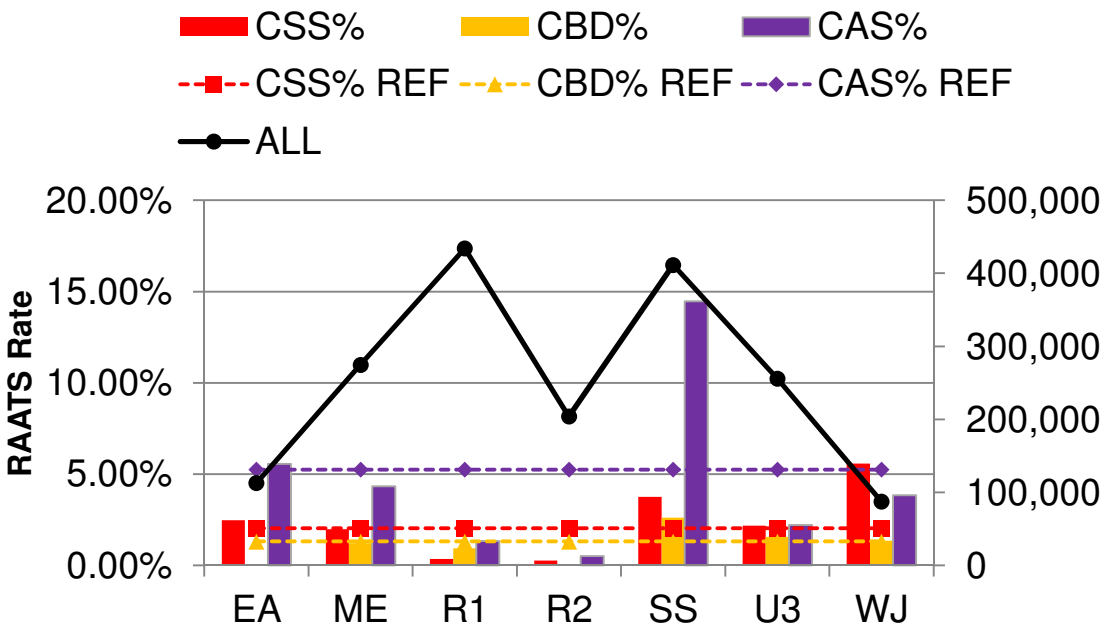
RAATS Tool User Interface

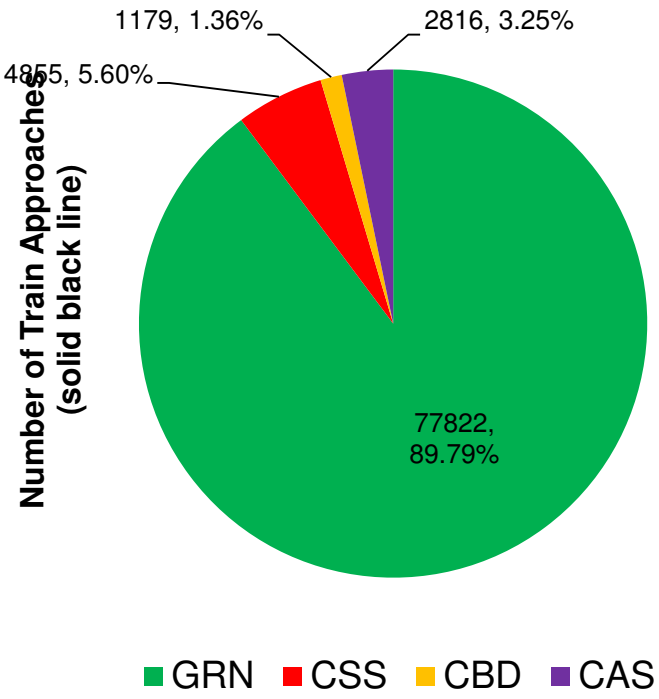
University of HUDDERSFIELD
Institute of Railway Research



Case Study – 7 TD Areas

University of HUDDERSFIELD
Institute of Railway Research



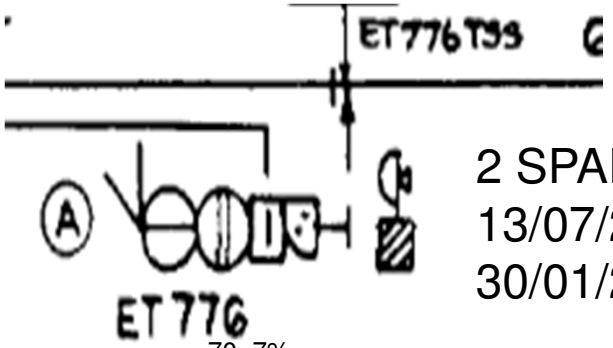


- 7 TD areas, 3 months data
- ~ 1.8M train approaches to 892 signa

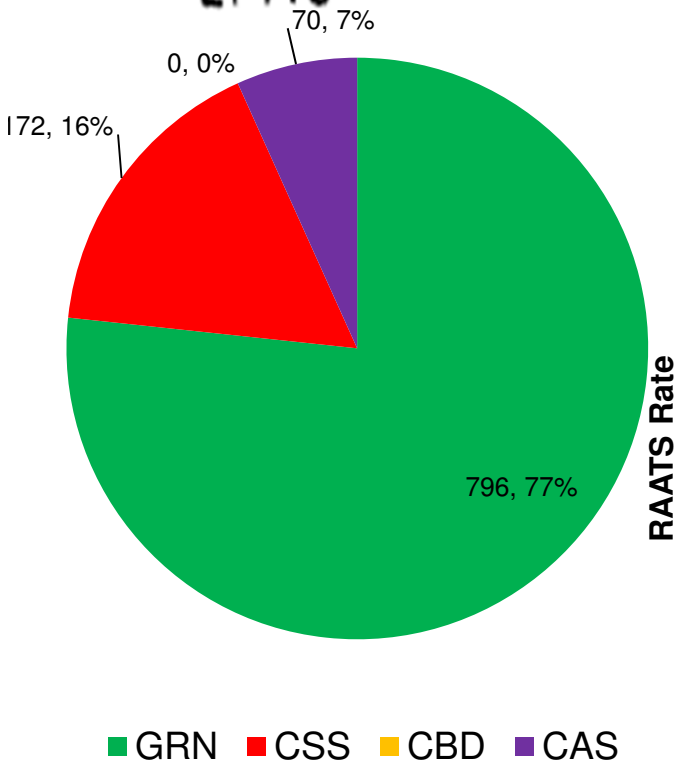
TD	Name
EA	Edinburgh IECC A
ME	Marylebone IECC
R1	Rugby SCC (Watford – Bletchley ITD)
R2	Rugby SCC (Northampton – Rugby ITD)
SS	Merseyrail
U3	Upminster 3 IECC
WJ	Watford Junction

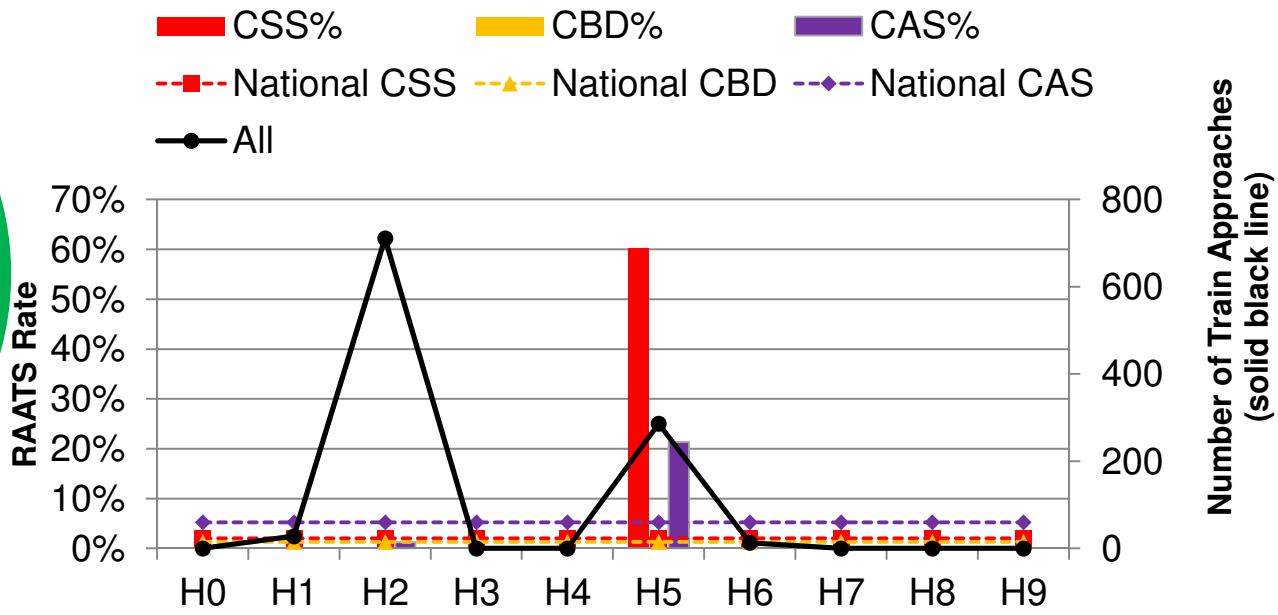
Case Study – Multi SPAD 1

University of HUDDERSFIELD
Institute of Railway Research



2 SPADs at
13/07/2011,
30/01/2014





Conclusions

University of HUDDERSFIELD
Institute of Railway Research

- An analysis model has been developed which allows the number of red signal approaches to be determined:
  - Over whole TD area or many TD areas
  - At individual signals or groups of signals
- The tool will allow:
  - Better understanding of SPAD risk at individual signals
  - Improved normalisation of SPAD data for trending and benchmarking
- An initial study of 7 TD areas showed that on average ~3.3% of signal approaches are to red signals
- There is a large variation between TD areas and between individual signals

Next Steps / Opportunities

University of HUDDERSFIELD
Institute of Railway Research

- **Validating the analysis model**
- Extending data availability from 7 to 61 TDs
- Improvements to the data presentation
- Studying the differences (similarities) between multi-SPAD signals
- Make the tool available via a web based interface?
- Consider using results as part of SORAT

Other opportunities:

- Info tailored to routes, TOCs, individual train services
- Timetable analysis / network operations
- Driver training
- …?

RAATS – BDRA Context

University of HUDDERSFIELD
Institute of Railway Research

- This is a useful stand-alone application of ‘newly available’ data for the industry

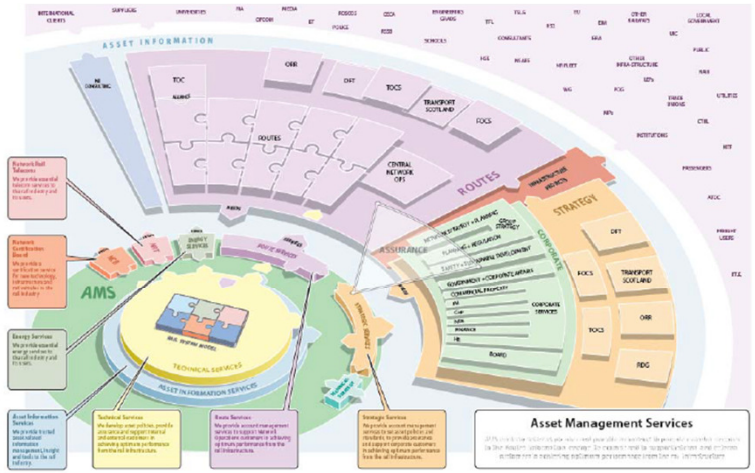
- In many senses it is still quite traditional
  - It only provides lagging information
  - It doesn’t blend several sources and types of data
  - It uses traditional IT software and infrastructure

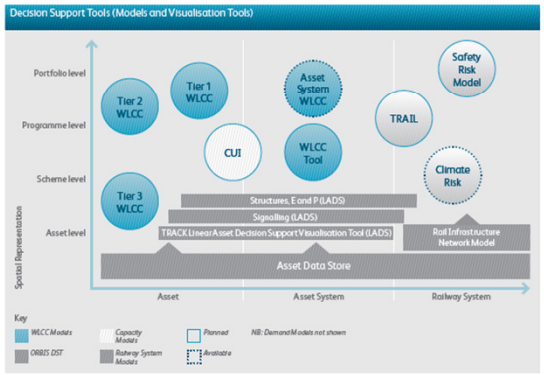
- However, it could play a part in ‘real’ BDRA by:
  - Being real time (or close to)
  - Being one source in the safety data picture
  - Using other big data to enhance accuracy or information

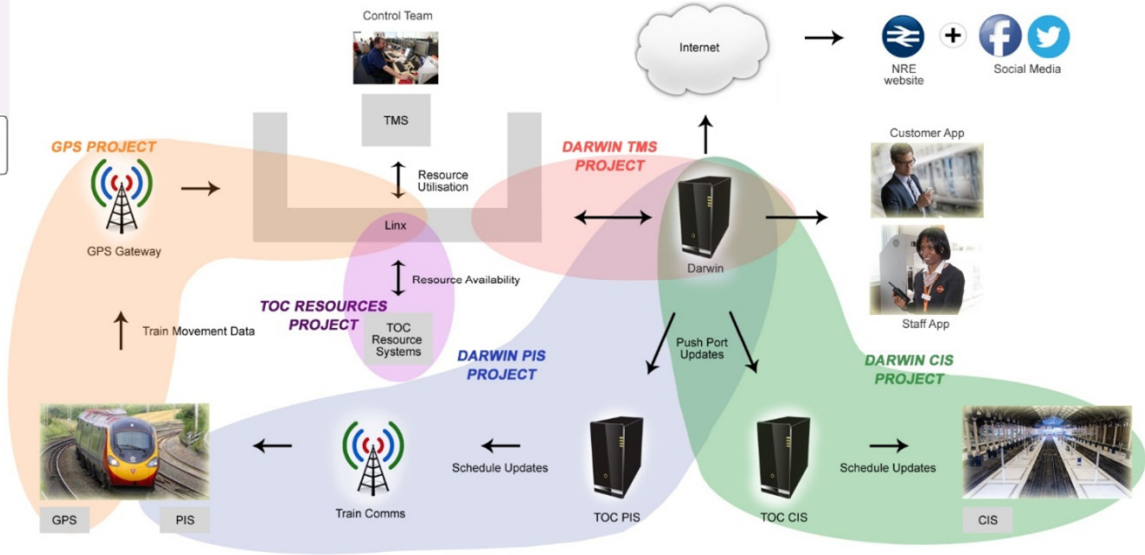
BDRA – Final Thoughts

University of HUDDERSFIELD
Institute of Railway Research

- We already have BD and more is on the way:
  - Network Rail Digital Railway
  - ATOC DARWIN







BDRA – Final Thoughts

University of HUDDERSFIELD
Institute of Railway Research

- We have to crack the RA part:
  - The key is how to make our data *useful*
  - How important is leading (vs lagging)?
  - How do we validate our results?
  - Will it allow us to do less, but more of the right things?
  - How will (could) BDRA make the railway safer?
  - Will it make the railway cheaper?

The only way of discovering the limits of the possible is to venture a little way past them into the impossible.

Arthur C. Clarke *Profiles of the Future: An Enquiry into the Limits of the Possible* (1962, rev. 1973), 21.

Hofstadter's Law: It always takes longer than you expect, even when you take into account Hofstadter's Law

Douglas Hofstadter *Gödel, Escher, Bach: An Eternal Golden Braid* (1979), 152

Acknowledgements

University of HUDDERSFIELD
Institute of Railway Research

The project has been undertaken under the Strategic Partnership between the University of Huddersfield and RSSB

The authors also gratefully acknowledge the technical support provided by Network Rail