

# Creating an Intelligent Railway System: from big data collection to information distribution

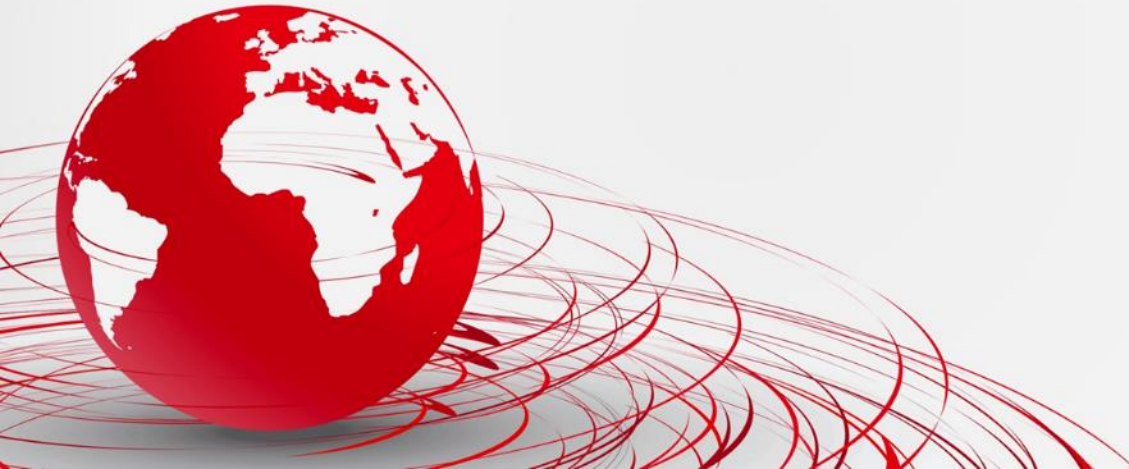
*24 November 2016*

*Big Data in Railway Operations*

*Intelligent Rail Summit 2016*

**Pietro Pace**

**MER MEC SpA**



# About MERMEC

MERMEC main goal is to supply integrated solutions for railway infrastructure inspection, measuring, data analytics and decision support.



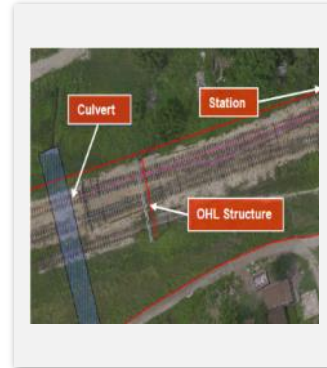
**MEASURING VEHICLES  
AND TRAINS**



**MEASURING SYSTEMS**



**DATA ANALYTICS &  
DECISION SUPPORT  
SOFTWARE**



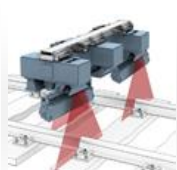
**MEASURING SERVICE &  
CONSULTANCY**

# Measuring systems and vehicles product portfolio

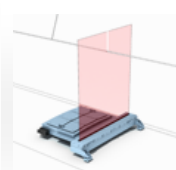
The wider and deeper portfolio on the market



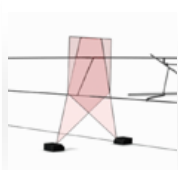
Track Measurement



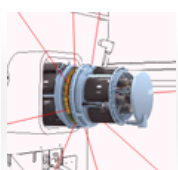
Track Inspection



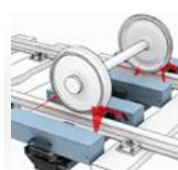
Catenary Measurement



Catenary Inspection



Tunnel Inspection



Train Monitoring



Measurement of Signalling & TLC

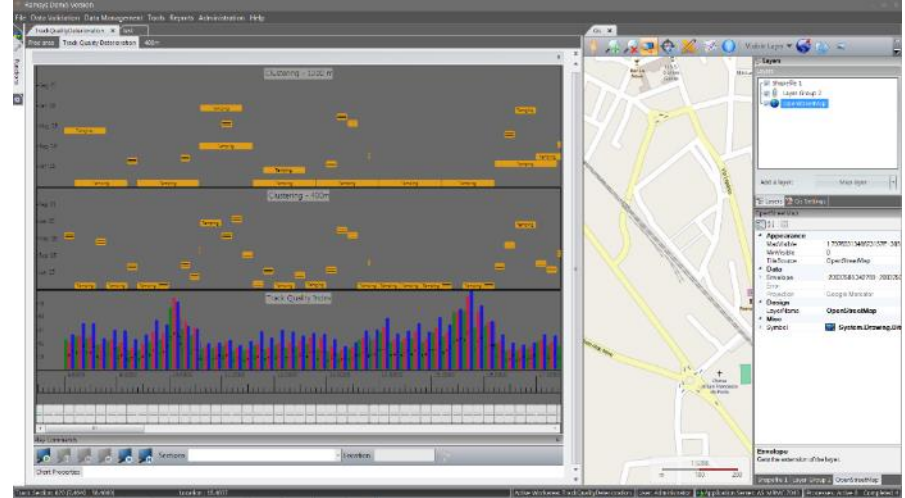
**60+ Types  
Measuring Systems**

**4 Models  
Recording Cars**

**1500+ Measuring and Inspection  
Systems Sold Internationally**



# New “big data” analytics and decision support



10+ TRACKWARE 20+ RAMSYS 10+ Countries

# Agenda

- 1 Multiple conditioning monitoring devices
- 2 Big Data Characteristics (7 Vs)
- 3 Manage big data
- 4 Accurately monitor and predict asset health
- 5 Extraction of new information (3 use cases)
- 6 Conclusions

# Agenda

- 1 Multiple conditioning monitoring devices
- 2 Big Data Characteristics (7 Vs)
- 3 Manage big data
- 4 Accurately monitor and predict asset health
- 5 Extraction of new information (3 use cases)
- 6 Conclusions

# Infrastructure Condition Monitoring

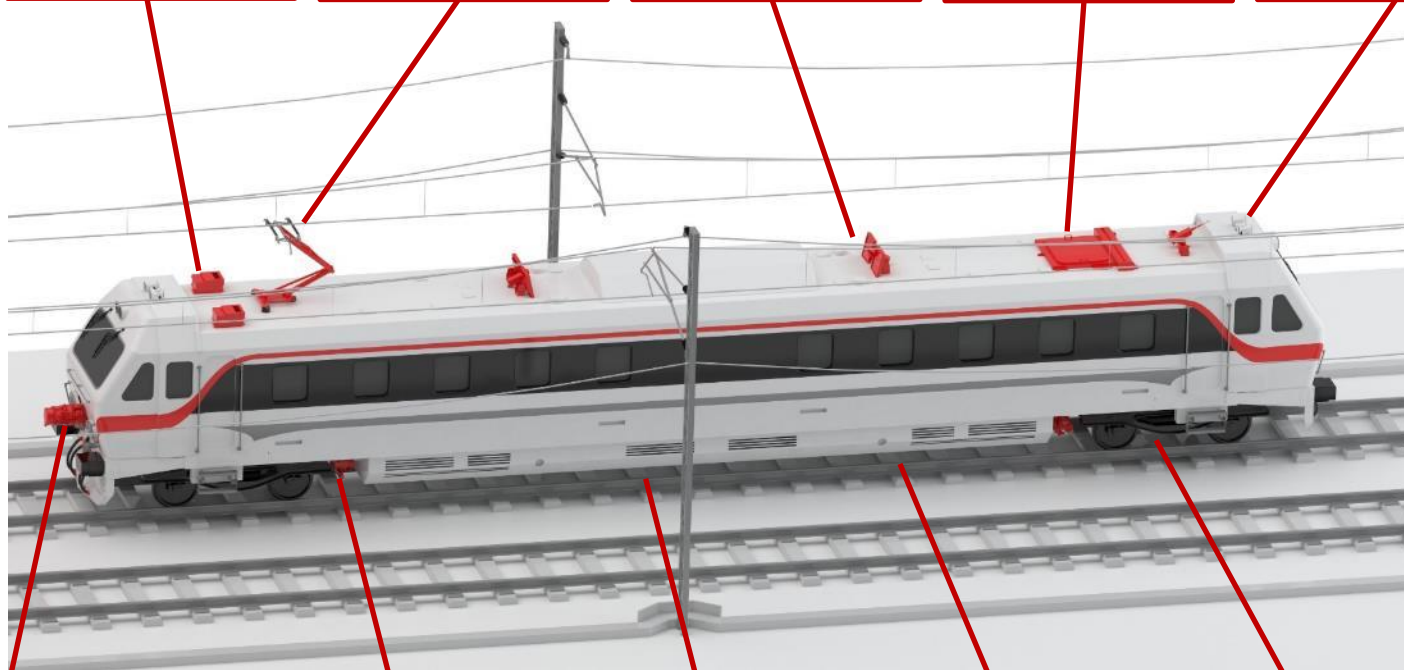
POLE & TUNNELS  
DETECTOR

PANTOGRAPH  
MEASUREMENTS

OHL INSPECTION

OHL GEOMETRY & WIRE  
WEAR

VIDEO SURVEILLANCE



CLEARANCE GAUGE &  
TUNNEL INSPECTION

TRACK GEOMETRY &  
RAIL PROFILE

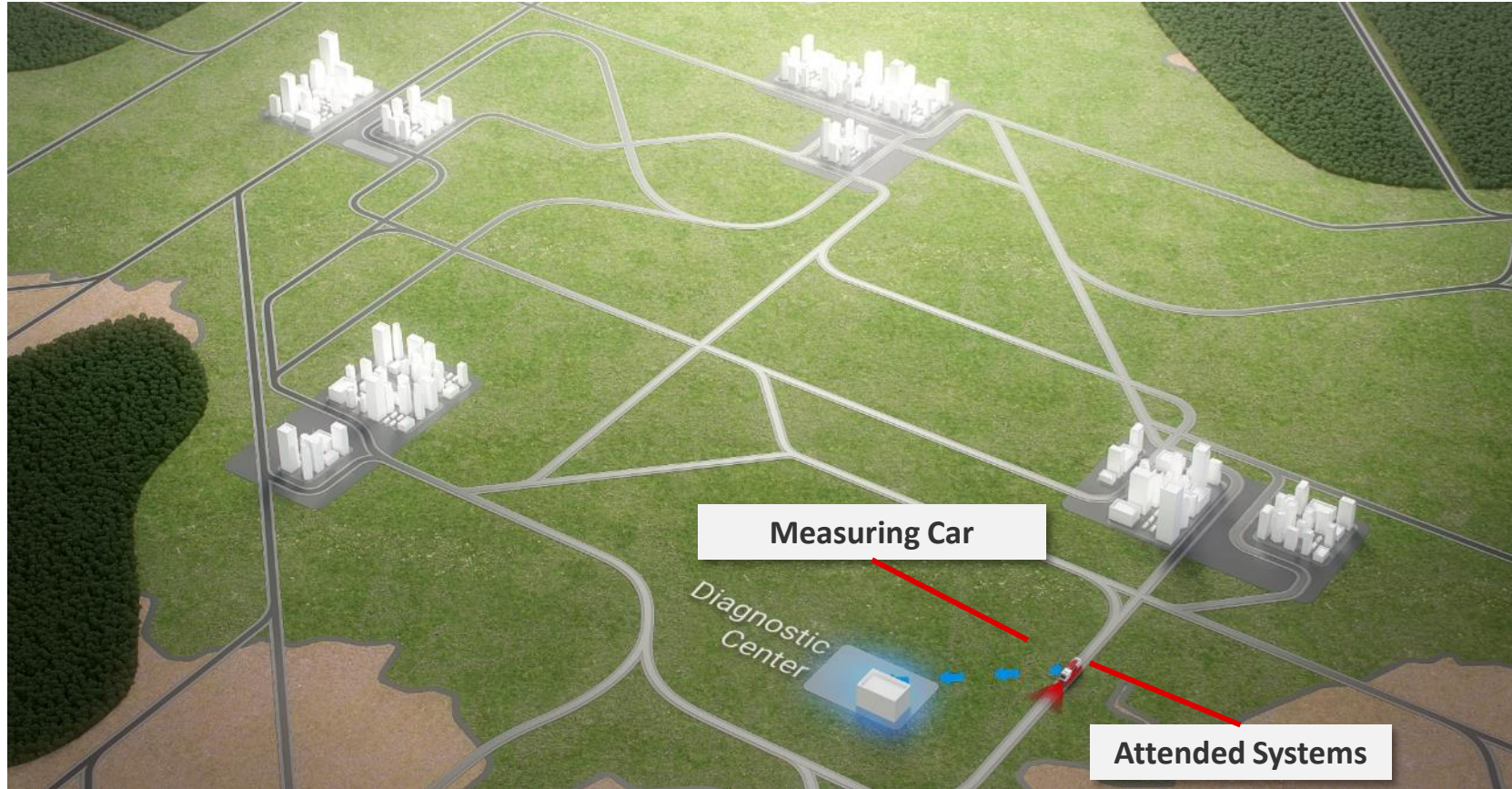
RAIL FLAW DETECTION

RAIL CORRUGATION

TRACK INSPECTION

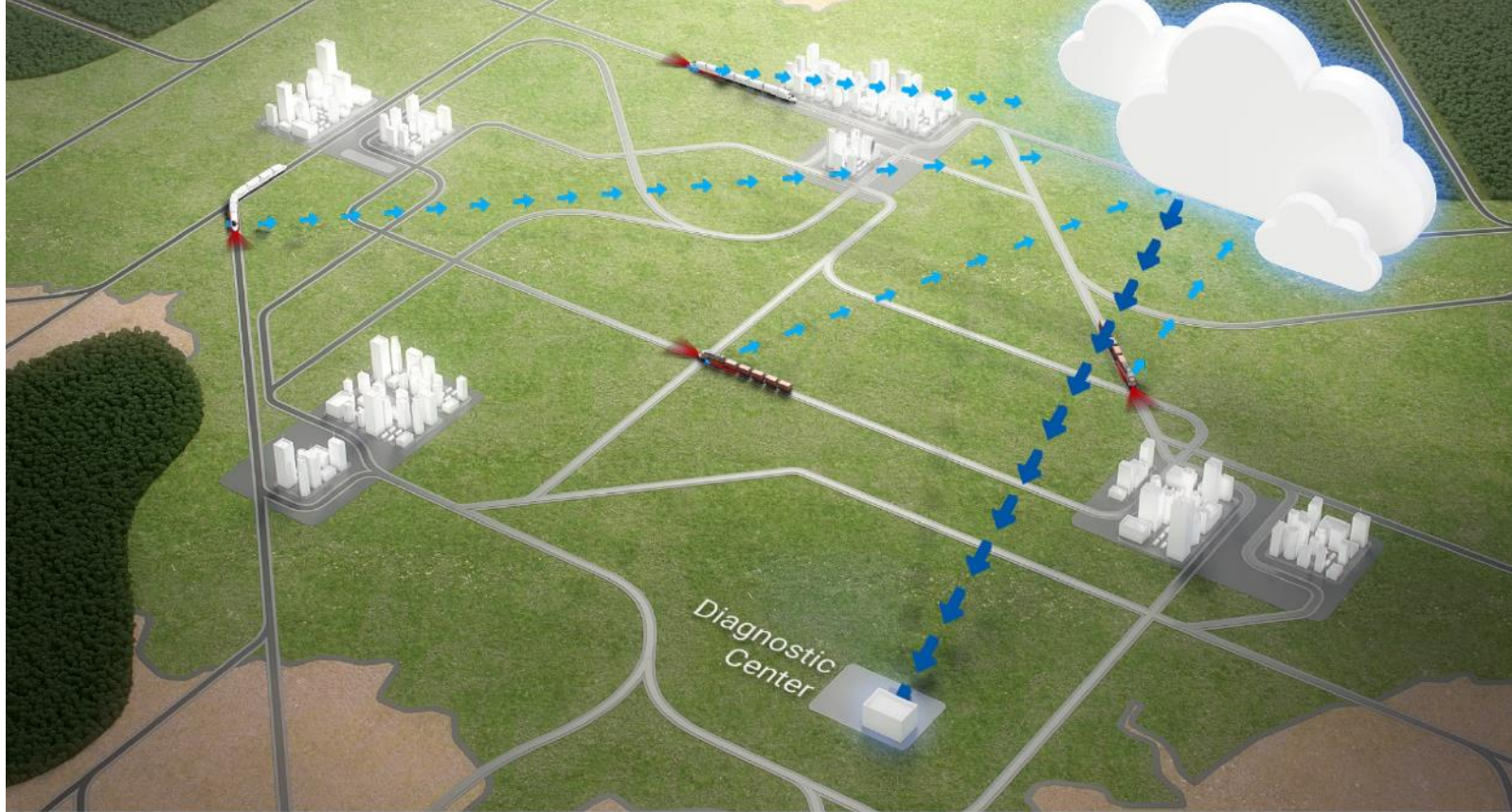


# Centralised Infrastructure Condition Monitoring

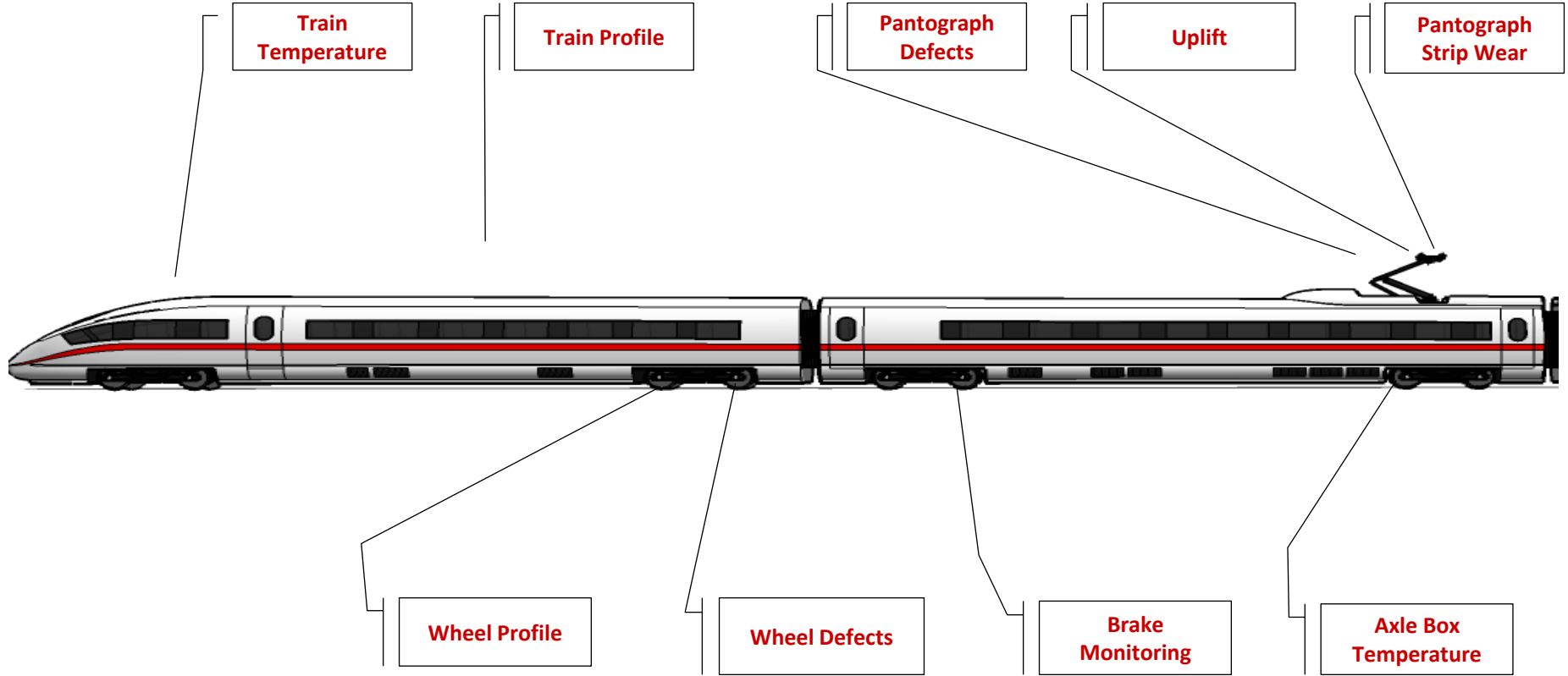




# Distributed Infrastructure Condition Monitoring



# Rolling Stock Condition Monitoring (Way Side)



# Lidar-Scanning Airplanes Surveys



1.

## LIDAR SURVEY

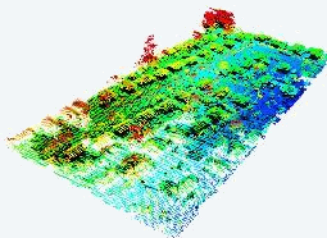
- LIDAR points
- Georeferenced Photos and Videos (aerial view)



2.

## DIAGNOSTIC VEHICLE SURVEY

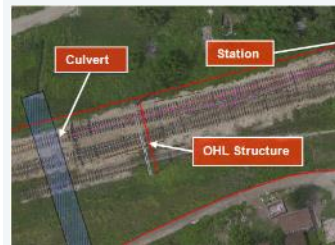
- Georeferenced Localisation
- Condition Data
- Photos and Videos (infrastructure view)



3.

## GEOGRAPHICAL DATA PROCESSING

- Corrected Geographical Data
- Digital Terrain Model
- Orthophotos



4.

## OBJECTS EXTRACTION

- GPS and Track Referenced Objects including Track centerline



5.

## ASSET CHARACTERISTICS POPULATION AND DATA TESTING

- Tested and validated Asset Data

# Agenda

- 1 Multiple conditioning monitoring devices
- 2 Big Data Characteristics (7 Vs)
- 3 Manage big data
- 4 Accurately monitor and predict asset health
- 5 Extraction of new information (3 use cases)
- 6 Conclusions

[illegible]

8 Bits = 1 Byte

1024 Kilobytes = 1 Megabyte

1024 Megabytes = 1 Gigabyte

1024 Gigabytes = 1 Terabyte

**1024 Terabytes = 1 Petabyte**

1024 Petabytes = 1 Exabyte

1024 Exabytes = 1 Zettabyte

1024 Zettabytes = 1 Yottabyte

1024 Yottabytes = 1 Brontobyte

1024 Brontobytes = 1 Geopbyte

1024 GeopBytes = 1 Saganbyte

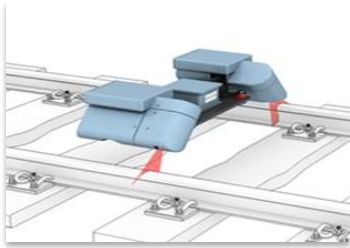
1024 Saganbytes = 1 Jotabyte

Year	Vehicle/Train Systems Configuration	GB/Mile	* Required Storage (TB)
1994	Track Geometry and Rail Profile	0.003	~0.5
2010	Several systems (no automatic inspection)	0.079	~14
2014	Several systems and automatic inspection	15.716	~2900

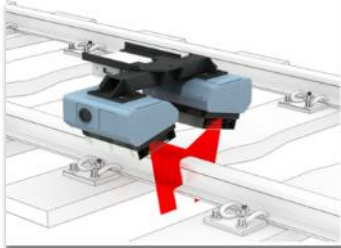




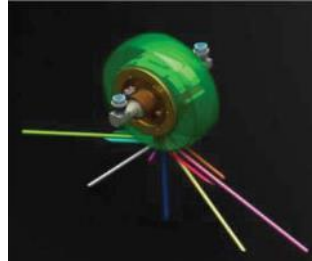
# Velocity: Inspecting and Measuring ...



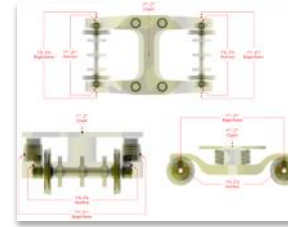
Track Geometry



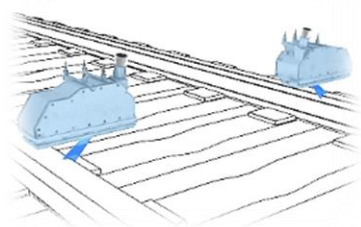
Rail Profile



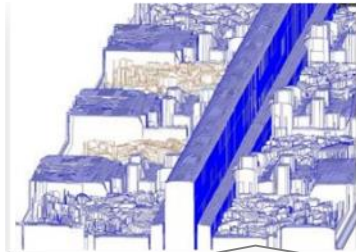
Ultrasonic Inspection



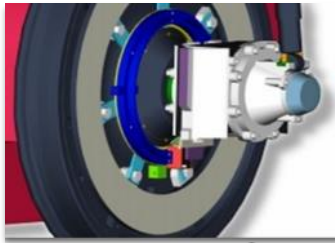
Running Dynamics



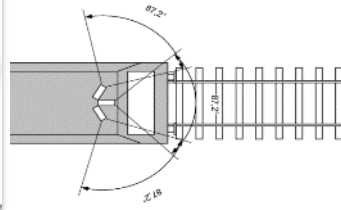
RCF Detection



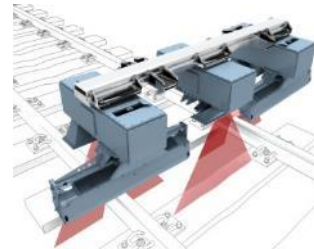
Ballast Measurement



Wheel/Rail Interaction



Track Video Surveillance



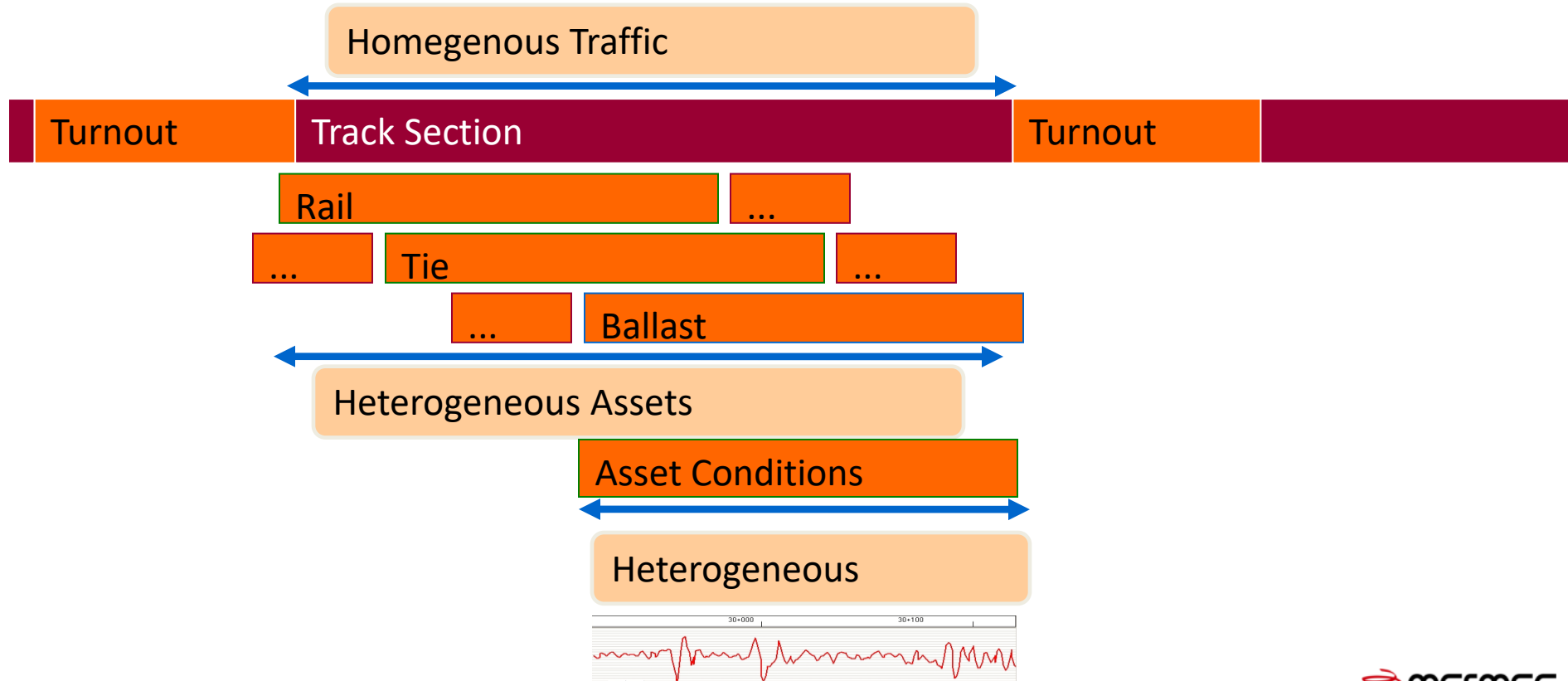
Vision Inspection



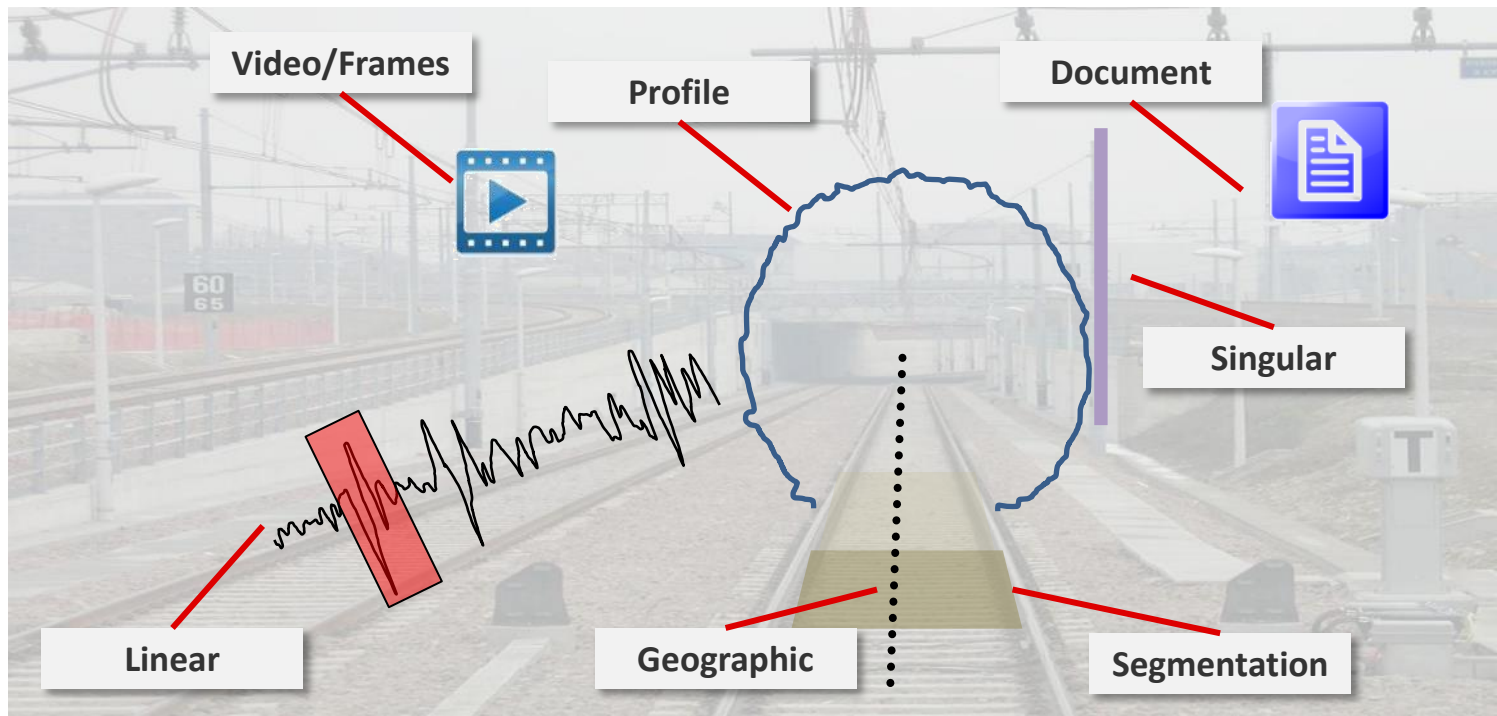
Tunnel

- **Measuring Speed:** Recording vehicle when hauled can reach 200 km/h
- **High Speed:** High Speed Recording Trains up to 400 km/h
- **High Frequency:** Autonomous/Unattended Systems (at each train run)

# Variability: Data at Different Levels

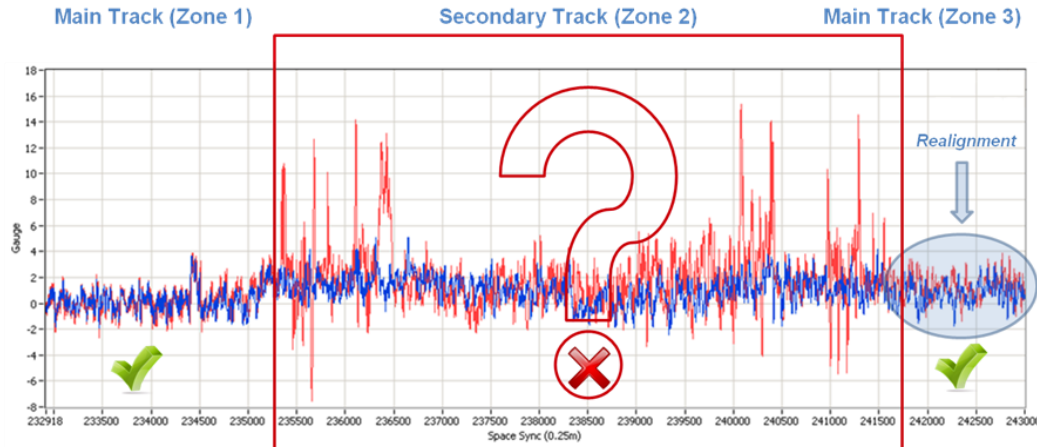


# Variety: Different Data Entity Types



# Veracity: Data Quality

- Diagnostic data are affected by typical uncertainty and in some cases by errors (e.g. Peaks, Noise, Miss-Calibrations, Wrong Master Data, etc.)
- Invalid data leads to invalid information (GIGO effect), in particular invalid defects/alerts so called **“False Positives”**
- **False Negatives/Gaps** can be generated by missed inspections (e.g. unavailable track, RCF, system failure, etc.)
- **“Unknown Positive or Negative”** for data that have invalid localization



# Visualisation: Common User Interface

False Positive

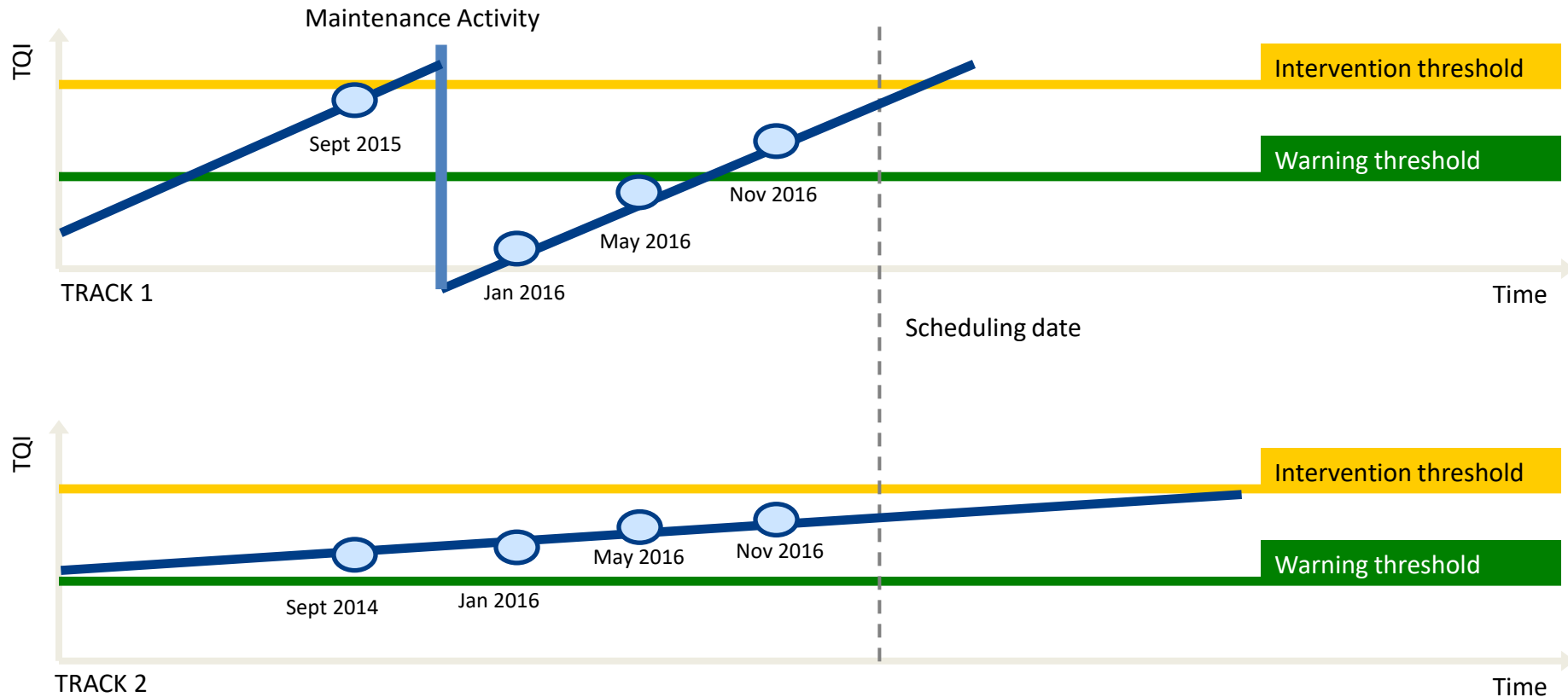
Valid Defect

Ballast

Track Geometry



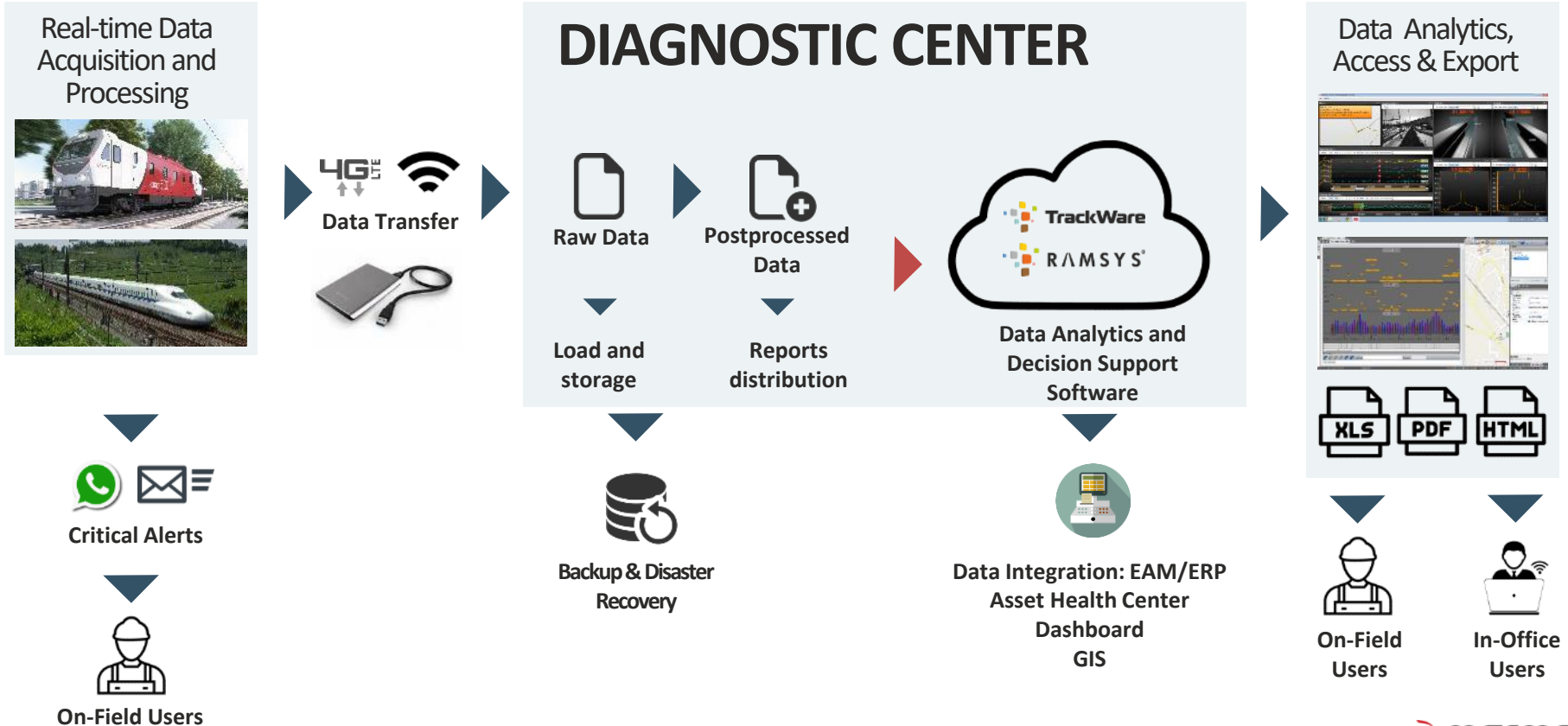
# Value: Why using big data?



# Agenda

- 1 Multiple conditioning monitoring devices
- 2 Big Data Characteristics (7 Vs)
- 3 Manage big data**
- 4 Accurately monitor and predict asset health
- 5 Extraction of new information (3 use cases)
- 6 Conclusions

# Managing Big Data and Data Integration



# Agenda

- 1 Multiple conditioning monitoring devices
- 2 Big Data Characteristics (7 Vs)
- 3 Manage big data
- 4** Accurately monitor and predict asset health
- 5 Extraction of new information (3 use cases)
- 6 Conclusions

# From data collection to data distribution

Data Import

Data  
Alignment

Measurement  
Validation

Data  
Processing

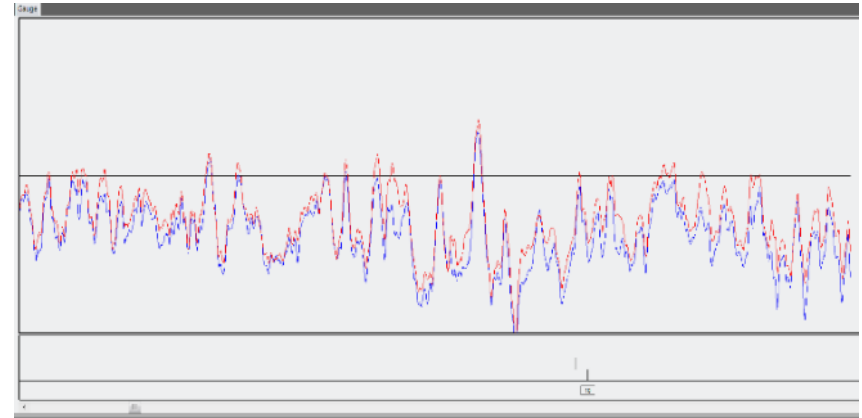
Defects  
Review

Data  
Distribution

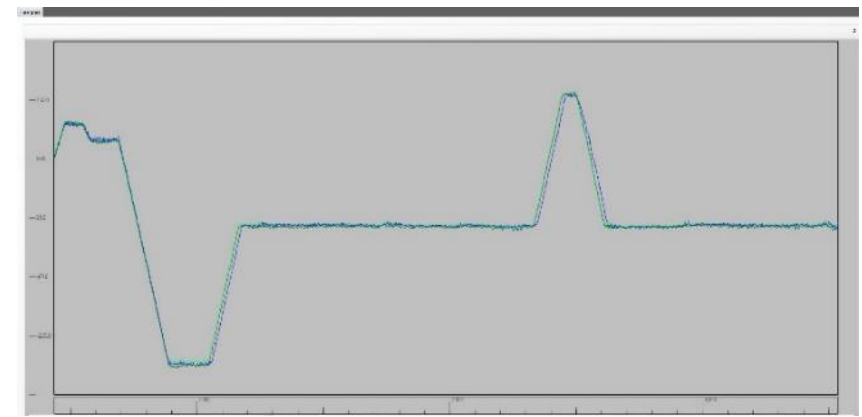
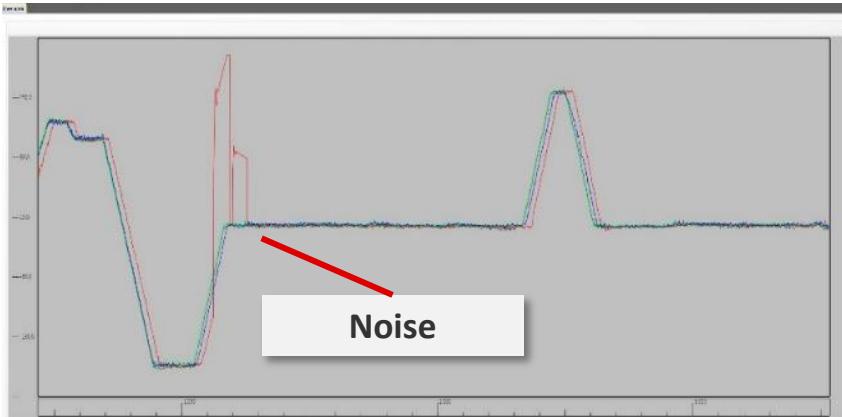


# Data Alignment and Validation

Position Shift



Noise



# Synchronise different data in space

Planned Works

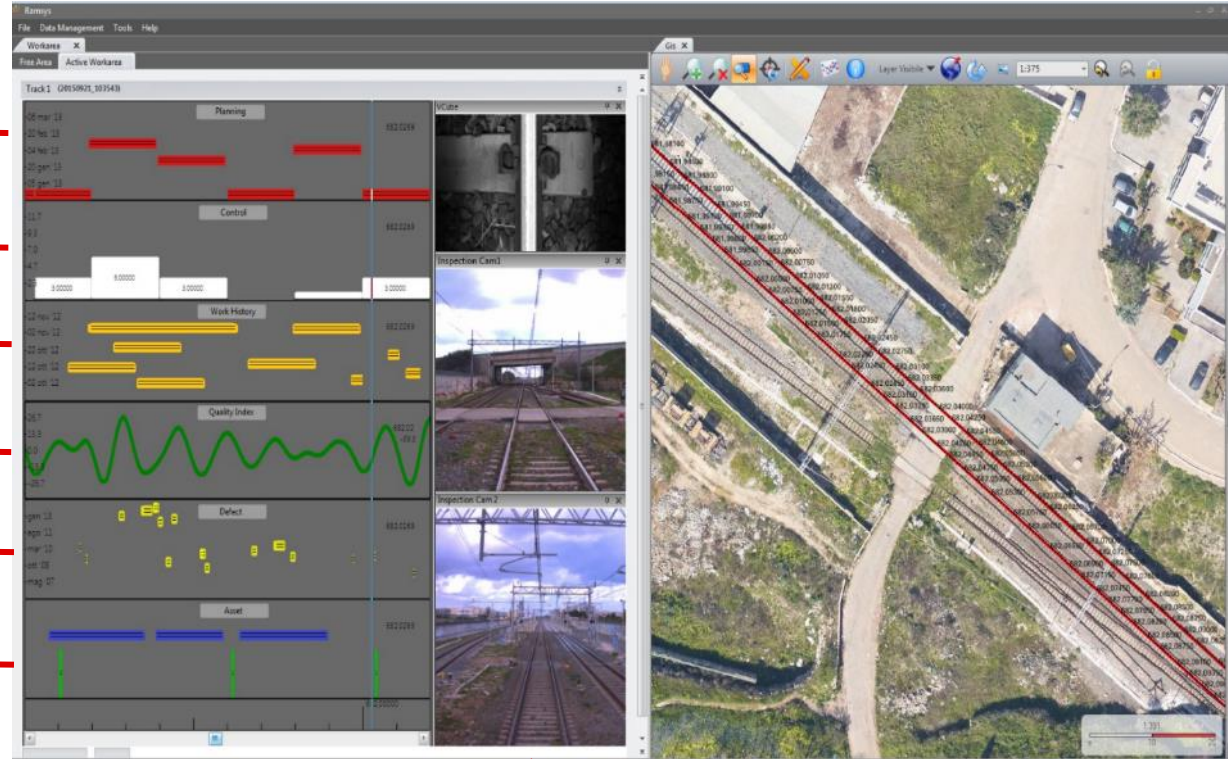
Key Performance Index/Score

Work History

Track Quality Index

Defect History

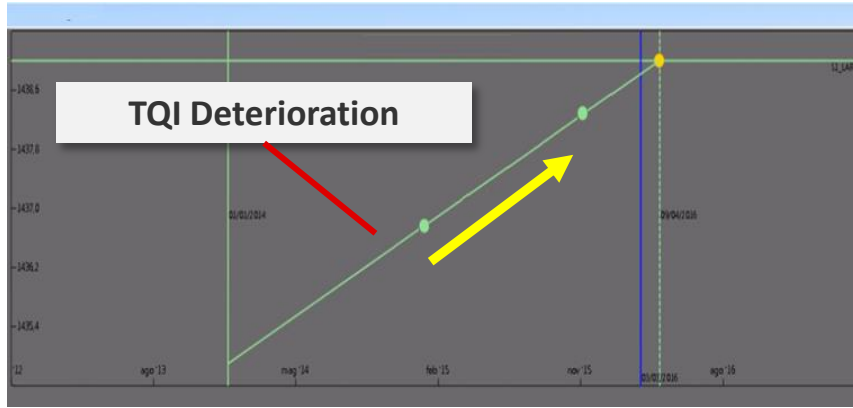
Infrastructure Features



Video Frames

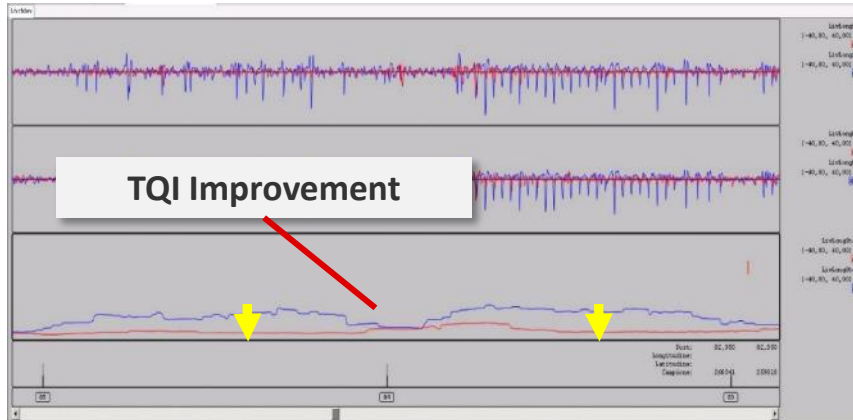
Geographical Data

# Process data in time

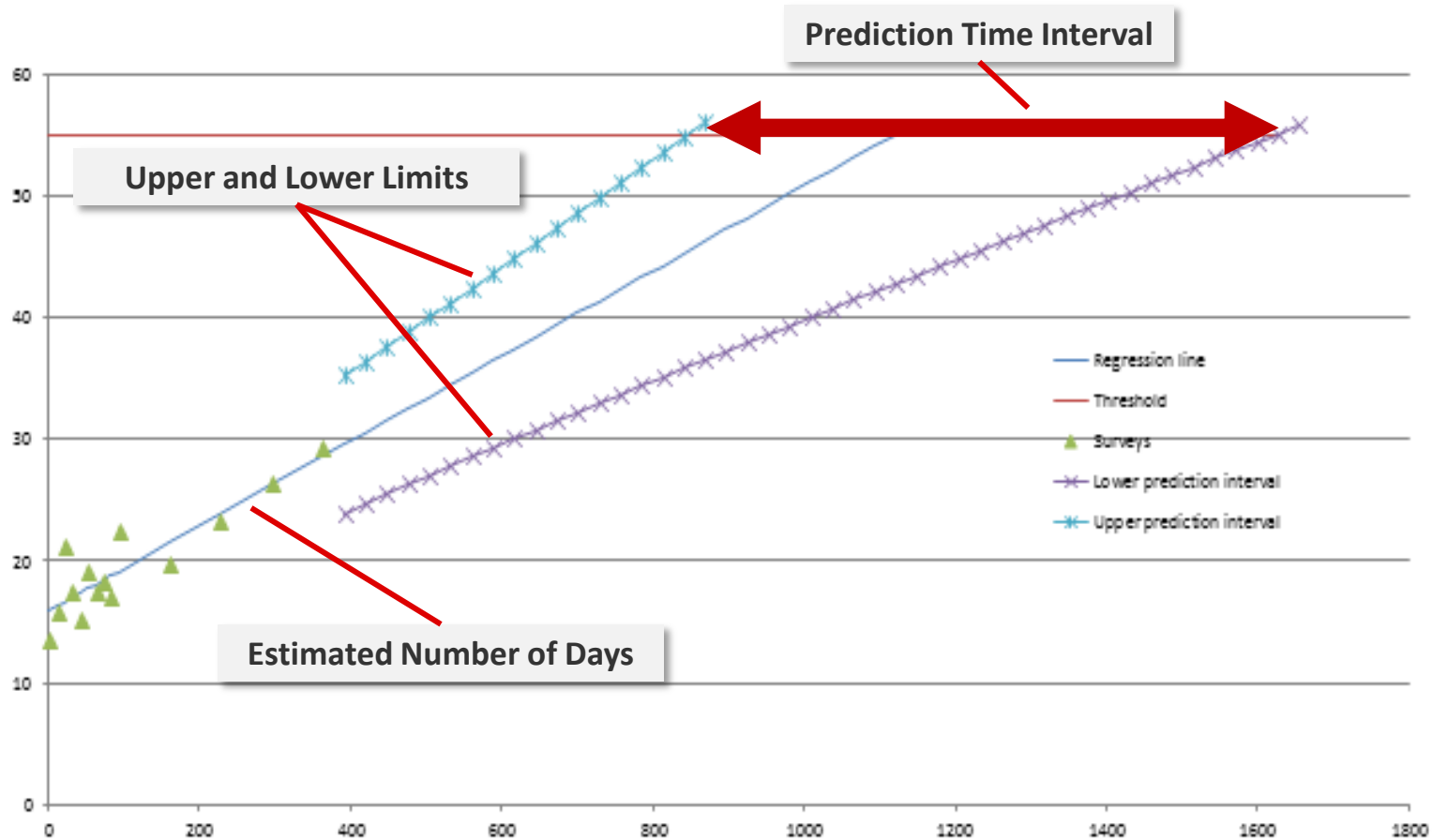


The rate of change can be calculated with only two Track Quality Indexes (TQIs).

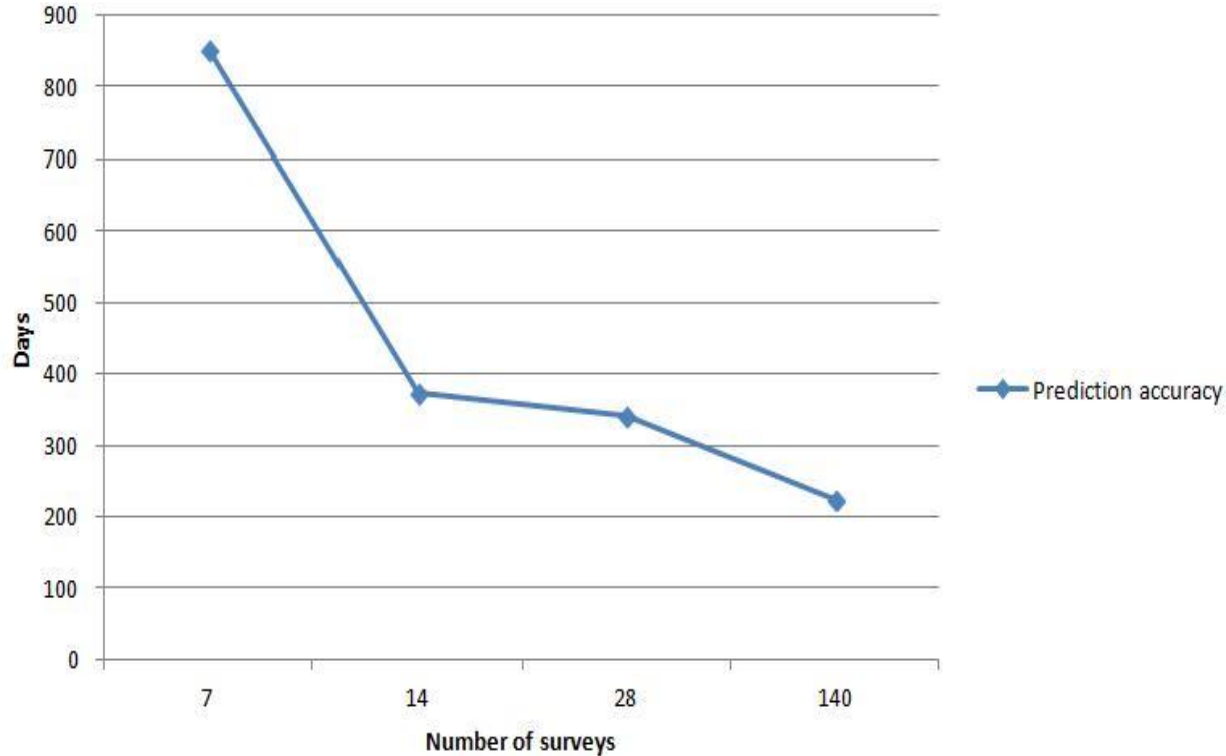
Accuracy of the rate of change calculated with only two TQIs?



# Prediction time interval



# Improved accuracy in using bigger data

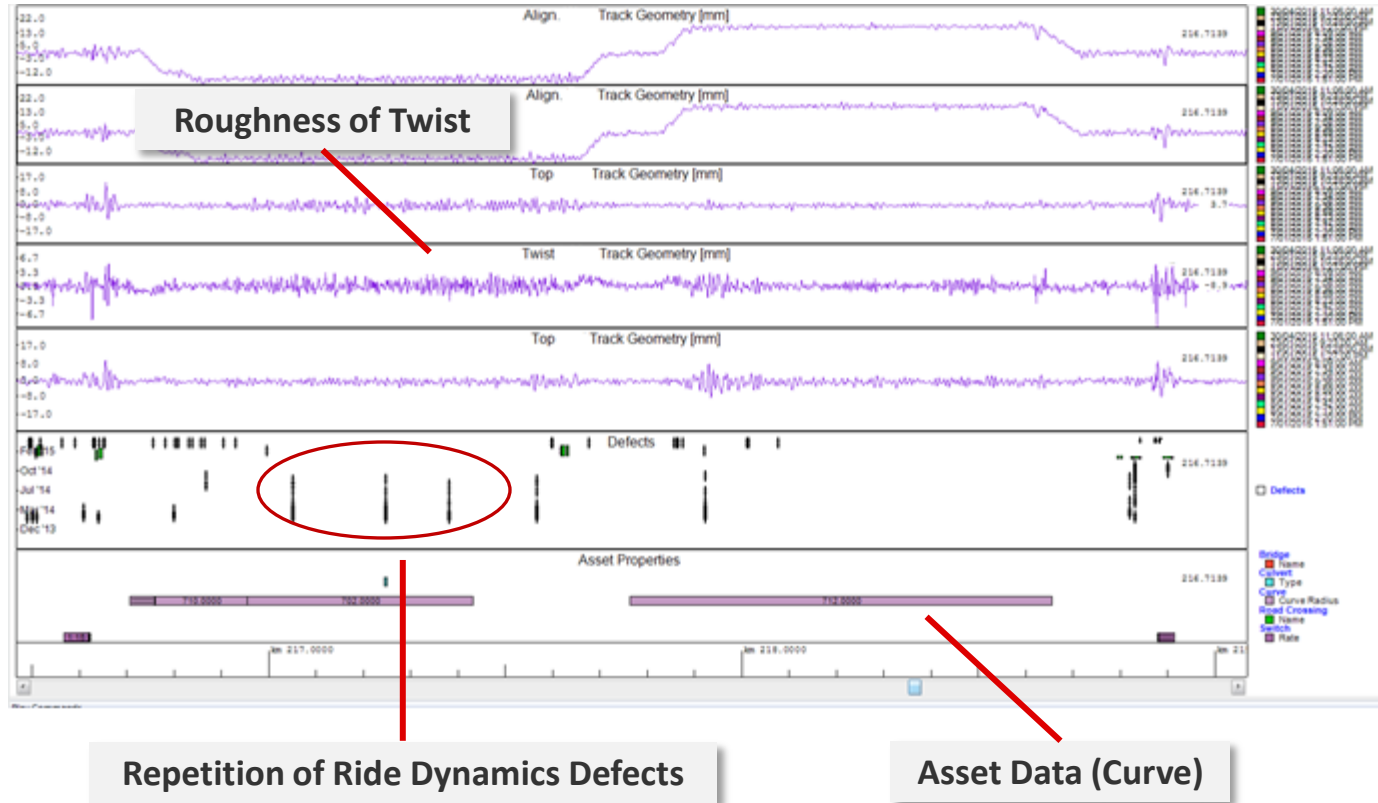




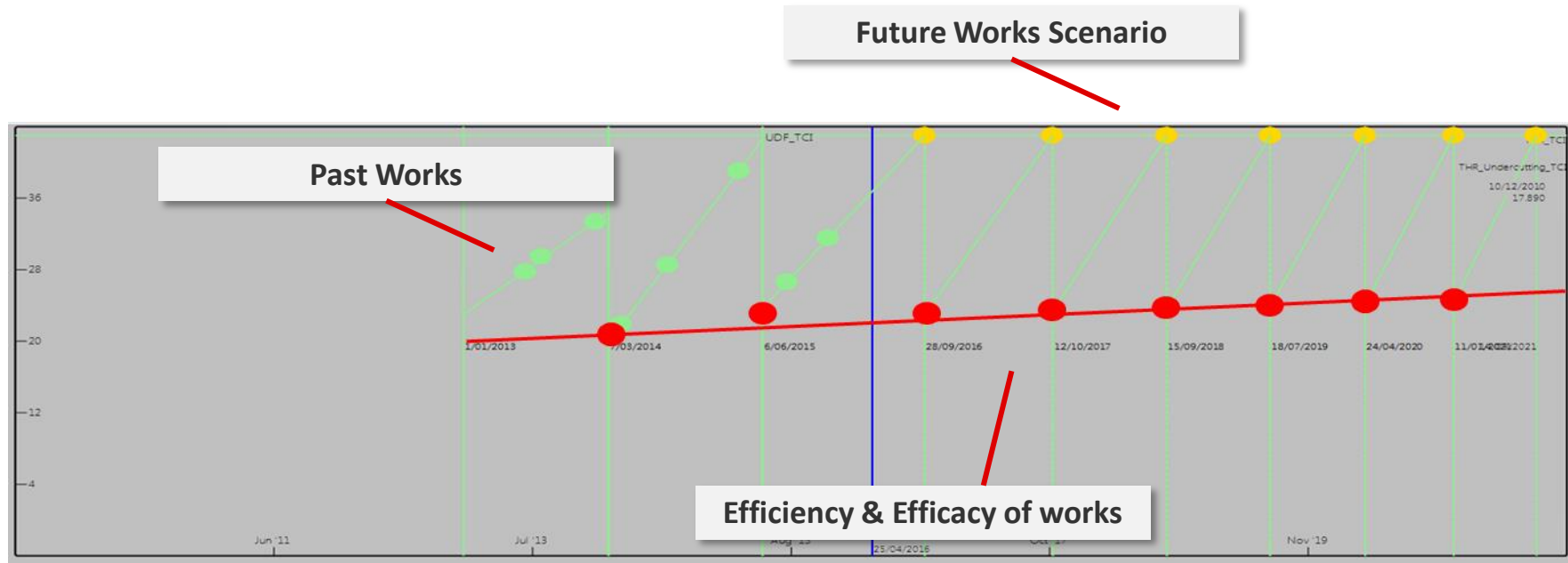
# Agenda

- 1 Multiple conditioning monitoring devices
- 2 Big Data Characteristics (7 Vs)
- 3 Manage big data
- 4 Accurately monitor and predict asset health
- 5** Extraction of new information (3 use cases)
- 6 Conclusions

# Correlate ride quality with track data



# Create scenario of maintenance & renewal works



# From prediction to prescription of a work plan

Ramsey's Demo Version

File Data Management Tools Reports Administration Help

Planning Report: HeadLoss Based

Code

HeadLoss Based

Description

Start Date

16/06/2008

End Date

16/06/2048

Workarea

Rerail East Tangent Tonnage vs HeadLoss

Template

Rerail East Tangent HeadLoss Based

Print Preview

Export to Excel

Chart

Show chart

Layout: <System>

Manage...

Row Filter

Drag a column header here to group by that column.

Type	Activity Date	Line	Track	Location From	Location To	Unit Cost	Fixed Cost	Total Cost	Actual Total Cost
rail East	02/05/2011			37.666	37.680	76400.000	0.000	1069.600	982.411
rail East	17/09/2019			37.680	37.685	76400.000	0.000	382.000	273.851
rail East	24/12/2020			37.685	37.708	76400.000	0.000	1757.200	1213.251
rail East	20/08/2035			37.722	37.738	76400.000	0.000	1222.400	844.001
rail East	25/03/2028			37.738	37.790	76400.000	0.000	3972.800	3160.642
rail East	25/03/2028			37.790	37.797	76400.000	0.000	534.800	369.730
rail East	21/12/2019			37.797	37.800	76400.000	0.000	229.200	163.050
rail East	21/12/2019			37.800	37.804	76400.000	0.000	305.600	217.401
rail East	10/07/2014			37.804	37.805	7640	0.000	76.400	63.855
rail East	09/05/2011			37.805	37.807	7640	0.000	152.800	140.265
rail East	28/03/2018			37.807	37.901	7640	0.000	7181.600	5801.579
rail East	26/10/2009			37.901	37.907	7640	0.000	458.400	325.705
rail East	31/03/2015			37.907	37.958	76400.000	0.000	3896.400	2768.490
rail East	05/01/2020			38.270	38.317	76400.000	0.000	3590.800	2551.353
rail East	27/03/2020			38.317	38.323	76400.000	0.000	458.400	323.551
rail East	04/10/2015			38.323	38.685	76400.000	0.000	27656.800	22286.258
rail East	27/12/2015			38.685	39.081	76400.000	0.000	30254.400	24214.174
rail East	14/05/2021			39.081	39.481	76400.000	0.000	30560.000	23812.388
rail East	10/04/2020			39.481	39.879	76400.000	0.000	437.827	312.526
rail East	18/08/2019			39.879	40.157	76400.000	0.000	5963.090	4610.110
rail East	31/08/2016			40.157	40.170	76400.000	0.000	9.103	6.103
rail East	25/05/2016			40.170	40.174	76400.000	0.000	61.635	43.235
rail East	16/06/2008			40.174	40.182	76400.000	0.000	611.200	483.269
rail East	05/07/2017			40.182	40.242	76400.000	0.000	360.295	260.119
rail East	31/03/2013			40.242	40.252	76400.000	0.000	63.102	45.102
rail East	11/02/2016			40.252	40.262	76400.000	0.000	609.111	459.111
rail East	11/02/2016			40.262	40.265	76400.000	0.000	40.265	29.265
rail East	28/05/2017			40.265	40.286	76400.000	0.000	40.286	29.286
rail East	28/04/2020			40.286	40.295	76400.000	0.000	484.061	344.061
rail East	28/05/2017			40.295	40.307	76400.000	0.000	611.200	459.111
Sum = 589886.702									

About 30% saving in life cycle costs

Total Actual Cost (NPV) = 590,000

Planning Report: Tonnage Based

Code

Tonnage Based

Description

Start Date

16/06/2008

End Date

16/06/2048

Workarea

Rerail East Tangent Tonnage vs HeadLoss

Template

Rerail East Tangent Tonnage Based

Print Preview

Export to Excel

Chart

Show chart

Layout: <System>

Manage...

Row Filter

Drag a column header here to group by that column.

Type	Activity Date	Line	Track	Location From	Location To	Unit Cost	Fixed Cost	Total Cost	Actual Total Cost
East	17/09/2019			37.680	37.685	76400.000	0.000	382.000	273.851
East	24/12/2020			37.685	37.708	76400.000	0.000	1757.200	1213.251
East	24/12/2020			37.708	37.722	76400.000	0.000	1069.600	738.501
East	24/12/2020			37.722	37.738	76400.000	0.000	1222.400	844.001
East	10/03/2016			37.738	37.790	76400.000	0.000	3972.800	3160.642
East	08/12/2020			37.790	37.797	76400.000	0.000	534.800	369.730
East	21/12/2019			37.797	37.800	76400.000	0.000	229.200	163.050
East	21/12/2019			37.800	37.804	76400.000	0.000	305.600	217.401
East					37.805	76400.000	0.000	76.400	63.855
East					37.807	76400.000	0.000	152.800	140.265
East					37.901	76400.000	0.000	7181.600	5801.579
East					37.907	76400.000	0.000	458.400	325.705
East	05/01/2020			37.907	37.958	76400.000	0.000	3896.400	2768.490
East	05/01/2020			38.270	38.317	76400.000	0.000	3590.800	2551.353
East	27/03/2020			38.317	38.323	76400.000	0.000	458.400	323.551
East	31/10/2014			38.323	38.685	76400.000	0.000	27656.800	22904.751
East	31/10/2014			38.685	39.081	76400.000	0.000	30254.400	25056.026
East	31/08/2016			39.081	39.481	76400.000	0.000	30560.000	23812.388
East	31/08/2016			39.481	39.879	76400.000	0.000	437.827	312.526
East	31/08/2016			39.879	40.157	76400.000	0.000	5963.090	4610.110
East	31/08/2016			40.157	40.170	76400.000	0.000	9.103	6.103
East	25/05/2016			40.170	40.174	76400.000	0.000	61.635	43.235
East	25/05/2016			40.174	40.182	76400.000	0.000	611.200	483.269
East	11/09/2013			40.182	40.242	76400.000	0.000	360.295	260.119
East	31/03/2013			40.242	40.252	76400.000	0.000	63.102	45.102
East	11/02/2016			40.252	40.262	76400.000	0.000	609.111	459.111
East	11/02/2016			40.262	40.265	76400.000	0.000	40.265	29.265
East	28/05/2017			40.265	40.286	76400.000	0.000	40.286	29.286
East	28/04/2020			40.286	40.295	76400.000	0.000	484.061	344.061
East	28/05/2017			40.295	40.307	76400.000	0.000	611.200	459.111
Sum = 840142.212									

Total Actual Cost (NPV) = 840,000

Ready

User: leonidas Application Server: localhost:6555

About 30% saving in life cycle costs

Total Actual  
Cost (NPV) =  
590,000

Total Actual  
Cost (NPV) =  
840,000

# Agenda

- 1 Multiple conditioning monitoring devices
- 2 Big Data Characteristics (7 Vs)
- 3 Manage big data
- 4 Accurately monitor and predict asset health
- 5 Extraction of new information (3 use cases)
- 6 Conclusions

# Conclusions

- ❑ Key characteristics of BIG DATA
- ❑ Key user requirements from current practices
- ❑ Big data collection (hardware) is not sufficient ....
- ❑ Two key questions for any railway:
  - (1) “How BIG is the decision support you are getting from current SMALL and BIG DATA?
  - (2) Are you sure you cannot make it even BIGGER?”



# Thanks for your kind attention!

"People who are really serious about  
**software**  
should make their own  
**hardware**"  
*Alan Kay*

