Railway Assets: a potential domain for big data analytics

Diego Galar

Professor of Condition Monitoring LTU
Head of Maintenance and Reliability Tecnalia
“Data is the new Oil. Data is just like crude. It’s valuable, but if unrefined it cannot really be used”

– Clive Humby, DunnHumby

“We have for the first time an economy based on a key resource [Information] that is not only renewable, but self-generating. Running out of it is not a problem, but drowning in it is.”

– John Naisbitt
Big Data ???

- 2010 = Structured, Homogeneous
- 2016 = Unstructured, Heterogeneous
Why Big Data in transportation? How We Got Here?
Meter Reads every 15 min.

M meter reads/month

B meter reads/day
Industrial and transportation data
a late but powerful entry

- 30 billion RFID tags today (1.3B in 2005)
- 4.6 billion camera phones worldwide
- 100s of millions of GPS enabled devices sold annually
- 2+ billion people on the Web by end 2011
- 12+ TBs of tweet data every day
- 25+ TBs of log data every day
- 76 million smart meters in 2009... 200M by 2014
- 2+ TBs of data every day
An increasingly sensor-enabled and instrumented business environment generates **HUGE** volumes of data with **MACHINE SPEED** characteristics...

**1 BILLION** lines of code
**EACH** engine generating **10 TB** every **30 minutes**!
SMART infrastructure and Decision Support Systems for maintenance in railway: Fusion of IM and Operators info.....

**On board data collection**

Condition indicators from real time data collection like:
- load
- vibration
- temperature

**Vehicle as moving sensor**

**Off line information**
provided by measurement vehicles

**On line information**

**DATA REPOSITORIES**

And

**INFORMATION FUSION**

**Track information**

Condition indicators from track side
SMART bearings as a sensor for condition monitoring

- Error detection of bearing
- Detect wheel damage
- Operation planning
- Position of the car
- Load in the car
- Maintenance planning
- Detect rail damage
- Continuous scanning of rail
- Error detection in boggie

LKAB

TRAFIKVERKET
Way side monitoring: enablers for big data services

Vibration, acoustic and strain gages
- Weight
- Rail forces
- Wheel damages
- Bearing damages
- Hunting, speed

Track geometry
- EMC
- Interoperability

Central server or cloud

- Financial information system (ERP)
- Geometry car data
- Field measurements and surveys
- Maintenance input records
- Field personnel
- Weather
- CMMS Maintenance database
- Track charts
Scale of Industrial Internet (of Things IIoT)

Social media versus electric generating power source

2012 Twitter Usage

80 Gigabytes per day
enabling social connections

Gas Turbine Compressor Blade Monitoring potential*

588 Gigabytes per day
enabling capital asset productivity

Data volume potential is 7x greater from a gas turbine than current Twitter usage

* Note: Assumes operational gas turbines (generating units only) >50MW are equipped with Blade Health Monitoring capabilities
Big Data is mostly machine generated data

Volume | Velocity | Variety | Variability

Machine-generated data is one of the fastest growing, most complex and most valuable segments of big data.
Industry 4.0, the concept

The basic principle of Industry 4.0 is that interconnected machines and systems will form an intelligent self-controlled network spanning the entire value chain. In the ideal factory, machines autonomously react to unexpected changes in production, predict failures and trigger maintenance processes.

How does this Industry 4.0 dream of technology-enabled just-in-time maintenance and near-zero downtime translate into reality?
The idea of Maintenance 4.0 and the connection with IoT
The IBM idea of Internet of things for maintenance

step one

connect securely to a new asset
The IBM idea of Internet of things for maintenance

**Step two**

Complete a process flow with the new streaming data
The IBM idea of Internet of things for maintenance

**Step three**

use predictive analytics to **forecast** asset failures
The need for analytics and sensemaking Big Data
Data must be prepared...
A fusion process which requires taxonomies and ontologies.
Taxonomy vs. Ontology

**Taxonomies:**
- Usually are a single, hierarchical classification within a subject
- Primarily focused on “isa” relationships between classes
- Limited in inferencing potential due to lack of relational expressiveness.

**Ontologies:**
- Subsume taxonomies.
- Include attributes with cardinality and restricted values.
- Unlimited relationships between entities.
- Superior inferencing support due to relational expressiveness.
Track model

- Failure data
- PM data
- Deployment locations
- Logistic parameter
- Cost data
The need of the context for sensemaking
Context-aware & CBM based Decision Support Solutions for maintenance actions

- Information models
- Knowledge models
- Context models

Data Fusion & Integration → Big Data Modelling & Analysis → Context sensing & adaptation

Link, Think & Reconfigure
Success Stories
To achieve a more effective way to conduct railway operation and maintenance
Measurement systems

Contact wire
Vision system
Track Logger
S&C Vision Logger
Boden
Luleå
Track Stability
Machine Vision System Inspection

Track Performance
Wheel Profile
Data providers and data owners:
- TRV
- JBV
- Measurement and other data from contractors
- LKAB
- Etc.

Project Data request

JVTC Management: Grants access to the data?

Manual

Automatic

New algorithms for analysis

JVTC: RRDC

Raw data

Data system for information logistic:
- Data quality check and cleaning, etc..
- Connecting data sets from different data providers, related to the same asset with the same Id.
  - Rolling stock and its components
  - Infrastructure and its components
- Relate data to other parameters like contextual information

Generates a new set of sorted data which will grant new possibilities for research and innovation.
Onboard monitoring. Impact from infra.

Maintenance contractors
Train operation

Wayside monitoring. Impact from traffic

Trafikverket
Infra. Managers

Maintenance contractors
Infrastructure
What do we expect from the analytics?
Prescriptive maintenance, one step ahead of prediction
Let us perform O&M according to our wishes and do not be passive watchers.
CBM++

- Sensing
- Measurement
- Diagnostics of Faults-Failures
- Prognostics
- Context aware RUL
- Decision Support Models
**Track/Train Maintenance Contractors**

**SCADA** Supervisory Control And Data Acquisition  
**ERP** Enterprise Resource Planning  
**CMMS** Computer Maintenance Management Software

**Data Collection**  
- Contactor Raw Data
- Sensors

**Signal Processing**  
- Transformed Data

**Feature Extraction**  
- Condition Indicators

**Fault Classification**  
- Diagnostic Modelling
- Failure Detection and Evaluation
- Contractual Condition Evaluation

**Managerial data**
- CMMS
- ERP
- SCADA

**Historical data**

**Required functions**
- Historical data
- Fault Classification

**Decision support system**  
- Remaining useful life  
- Reliability prediction  
- Risk assessment

**Prognostic Modelling**
Prognosis, enabler for new business models

- Work Orders/CMMS
- Physics of failure
  - Pavement degradation, track geometry, ballast aging etc..
- On board collected data from vehicles
- IM and construction data
- Data driven methods
  - Condition indicators from infrastructure: ON LINE+ OFF LINE
- Most restrictive state condition assessment
- Accurate diagnosis
- RUL estimation
- Most restrictive RUL estimation
- IM decisions on maintenance
- O&M decision making
- End user decisions on new services
- Prognosis, enabler for new business models

Symbolic description, and experience based information
Data science... Narrow vision and mistakes
Value when analyzing data at mass scale

• As observations increase in frequency
  – Each individual observation is worth less
  – ...as the set of all observations becomes more valuable

• *Big Data* is the accumulation and analytical processes that uses this data for business value
Value when analyzing data at mass scale

• As observations increase in frequency
  – Each individual observation is worth less
  – ...as the set of all observations becomes more valuable

• *Big Data* is the accumulation and analytical processes that uses this data for business value
Let us be careful bigger = smarter?

• Yes!
  – tolerate errors
  – **discover the long tail and corner cases**
  – machine learning works much better

• But!
  – more data, more error (e.g., semantic heterogeneity)
  – with enough data you can prove anything
  – still need humans to ask right questions, lack of analytics
But Remember...

Not everything that can be counted… counts,
Not everything that counts… can be counted

The only thing that interferes with my ability to learn is…

My Education

AND…even with all this data

We can’t find many answers
Black Swan Losses

• Loss Distribution
  – Tail events are rare – very little data
  – Typically strong model assumptions
Why?
Data driven or model based?

Data-Based or Physics-Based Models? – That is the question!
Hybrid & Context Driven Services

- Physics of failure based
- Hybrid models
- Data driven
- Context Driven Services
- Context Awareness
**Track/Train Maintenance Contractors**

**Data Collection**
- Contactor Raw Data

**Signal Processing**
- Transformed Data

**Feature Extraction**
- Condition Indicators
  - DATA FUSION
  - Diagnostic Modelling

**Fault Classification**
- Fault Detection and Evaluation

**Managerial data**
- CMMS, ERP, SCADA

**Historical data**
- DATA FUSION

**Required functions**
- Contractual Condition Evaluation

**Data Fusion**
- Decision support system
- Remaining useful life
- Reliability prediction
- Risk assessment

**Prognostic Modelling**

**Track/Train Maintenance Contractors**

**SCADA** Supervisory Control And Data Acquisition
**ERP** Enterprise Resource Planning
**CMMS** Computer Maintenance Management Software
**Track/Train Maintenance Contractors**

**SCADA** Supervisory Control And Data Acquisition
**ERP** Enterprise Resource Planning
**CMMS** Computer Maintenance Management Software

**Data Collection**
- Contactor Raw Data

**Signal Processing**
- Transformed Data

**Feature Extraction**
- Condition Indicators

**Fault Classification**
- Diagnostic Modelling
- Contractual Condition Evaluation

**Synthetic Data**
- Physics of failure

**Data Fusion**

**Decision Support System**
- Remaining useful life
- Reliability prediction
- Risk assessment

**Data Fusion**

**Train/Track Contractor**

**Sensors**

**Work Order**
Concluding remarks

• Transportation data is becoming one of the largest domain for big data analytics and target of data science
• Massive observations of certain processes do not assure the quality of deliveries
• Lack of definition in the services expected by customers