# **Yellow Trains (YT): Geometric Data Analysis**

# **Challenge Statement**

Network Rail (NR) and the Connected Places Catapult are publishing this notice to explore the state of the market in relation to the business challenge described below.

The purpose of this notice is to provide potential suppliers with an overview of the challenge and seek responses that explain the capabilities of products and solutions available in the marketplace, or near-market prototypes. A variety of potential solutions are sought to this challenge, and there is no presumption that respondents will have previous experience operating in a rail environment.

Please respond by completing the template attached to this notice.

## **Context**

NR rail has a fleet of 7 dedicated track geometry recording trains, known as the Track Geometry Infrastructure Monitoring Fleet, (aka Yellow Trains (YT)). These trains traverse the rail network and scan the infrastructure to monitor the condition of the infrastructure and report on its status. The YT fleet is equipped with a range of measurement technologies that can be configured in different ways, depending on the type of infrastructure being recorded.

The YT fleet operate from Derby and are dispatched on scanning activities based on plans that are committed to up to 3 years in advance. Each shift includes specific sections of the track that must be recorded. Sections of track are recorded as defined by the Track Geometry train recording plan, a reference file called Journey Files or Route Setting Tapes are used to identify and record specific sections.

Based on short term needs, the scanning activities can be changed up to a week ahead based on operational needs.

The YT fleet have technicians on board that monitor the track geometry data being collected in real time. Critical faults are classified as “fix within 36hrs” or in severe cases the line can be blocked to traffic. YT On Train technicians report faults that are classified as high priority directly to the Signal and Control Centre and to Track Maintenance Engineers. Track engineering responsibility on the network is broken down into Routes and Regions. The Routes are:

* London North-eastern
* London North-western
* Midlands & Continental
* Anglia
* Western
* Wales
* Scotland
* Wessex
* Sussex
* Kent

Multiple data streams are collected on the YT fleet, but for the purpose of this notice, we will be focusing on a solution that supports the Track Geometry (TG) team.

Currently, the TG team, which is part of the Asset Information Services (AIS) team at NR, is responsible for processing and reporting of track geometry data to the track engineers. The TG data consists the following primary track parameters.

* Twist over 3m
* Dip Left and Right in mrads
* Top Left and Right rail 35m
* Alignment 35m
* Mean Top 70m
* Mean Alignment 70m
* Gauge
* Cross level
* Curvature
* Cant Deficiency

Data processing activities within the team is not fully optimised and consists of automated process, semi-automated process and manual process of track geometry data.

The whole end to end process that the TG staff need to follow is comprised of multiple tasks and sub tasks involving cross checking of information copying and conversion of data into new formats.

Some tasks are manual and time consuming in nature. Involving the use of multiple software tools, paper-based information and cross comparison between multiple windows, file storage locations (Server and local C: drive based) and email communications (which include .pdf files).

All the tasks are essential for the successful delivery of TG data and to support a safe rail network infrastructure.

This complex process requires a lot of experience and tacit knowledge to generate the correct outputs and is prone to human error.

The ambition of this PIN is to identify an automated solution that will tackle the challenges outlined in this document to improve the efficiency of the TG team’s data processing role, resulting in a significant reduction in data processing time.

There are many technologies which may assist in automating the TG data analysis processes, including Software Robots (RPA), AI, ML and data integration. This PIN aims to identify potential suppliers to NR that can bring forward technologies and strategies that are new and novel and could potentially expand into other areas of data processing within the NR organisation.

**Data Analysis Status Quo**

Track Geometry Data is recorded by 7 track recording vehicles (TRVs) that have track geometry equipment on board. The network is broken down by ELR (Engineers Line Reference), Track ID and Mileages. Recording of these sections is based on the infrastructure Monitoring train plan. TRV’s use reference files such as Route Setting Tapes (RST’s) or Journey files to control and manage the recording and delivery of this data.

Recorded information is sent to a NR server and an email is sent to the TG team to notify them of a completed shift in the form of a logsheet. A summary of data the collected is contained within a Log Sheet that contains multiple RSTs

The Track geometry Team then undertake their data analysis process. See Figure 1. below.

At the end of the data analysis process, fault reports, trace, updated logsheet and invalidated faults spreadsheet are compiled and sent to a track engineer responsible for a section of network. Data is also uploaded into downstream data repositories/ portals such as Track Geometry Reports (TGR), LADS and TIGER databases, which electronically stores track geometry data. The responsibility of these databases does not fall under the track geometry team.

#### **Track geometry data analysis process**



Figure 1 Existing geometry data analysis stages

As seen from the diagram in previous page the process has 6 stages:

* Data Receipt
* Data Archiving and conversion.
* Data Integrity Check
* Run on Run Check
* CDMS
* Data Distribution and TGR/TIGER/LADS upload

All 6 data analysis stages are summarised in more detail below.

#### **Data receipt stage summary**

|  |  |
| --- | --- |
| Stage | Actions |
|  | Data collection train (yellow train) collects the TG data (along with other data files not used by TG team). The On Train Technicians emails the logsheet to track geometry team email address and the TRV automatically uploads data on the server. The logsheet is emailed after the shift is complete, but data is automatically transferred over after each run. Possible delays in data receipt due to the incorrect process of power shut off on the train or network connectivity issues.  Data format and file from all trains is consistent.  Process improvement ideas/notes:   * Common graphical interface use for all data analysis processes * Notification in the interface not in the email * Team leader or analysts assign the task in the interface * Task progress can be tracked (by analysts or team leader) throughout the interface * KPI spreadsheet can be incorporated with live data analysis progress. |

Table 1 Track Geometry Data Receipt Stage

#### **Data archiving stage summary**

|  |  |
| --- | --- |
| Stage | Actions |
|  | Tracksys converter creates run on run file at the start of the cataloging process. Tracksys is an internal software that is run automatically and saves the files in the Tracksys folder.  Data archiving consists of the following tasks for each track run:   * folder creation * data extraction * unarchiving * sorting of data   Archiving stage uses VBA excel tool created by TG team. VBA file actions combine automatic and manual actions (which can be automated).  Archiving steps:   1. Log file check of what data was captured against the planned data capture (manual check). 2. Archiving tool inputs.  * Path of the folder where to catalog data (manual input) * Day of run (manual input) * Track recording vehicle (manual input) * Run number * Data recording * Route setting tape * Journey file number * Date before and after recording. * Removal of duplicate rows or lost runs (manual).  1. Archiving tool actions  * Run renaming * Folder creation * Route setting tape transfer from **C drive** to run folder * Journey file transfer from **masterstore location** (server) to run folder * TP file csv file trace file, geo file transfer from tracksys server to run folder * Clean table.  1. Manual extraction of 5 text files from the .zip file in the run folder.   Process improvement ideas/notes:   * RST files automatic update each 4 weeks, no need for manual update * All tasks at this stage can be automated by the interface, incorporating unarchiving of the data * Data transfer can be done with one click instead of multiple, if the aim is main result – cataloged data. * VBA stage needs to be completely automated. |

Table 2 Geometry Data Cataloguing Stage

#### **Data integrity check stage summary**

|  |  |
| --- | --- |
| Stage | Actions |
|  | Data integrity check uses the text files generated by the track recording vehicle and PCRST software interface for reading RST file. Comparison between RST and text files is done manually to check the geo location reference in text files. RST file is hard to interpret as the text file. PCRST is IT supported application that converts RST data into more readable format.  Also, VBA code is used for Microsoft access. VBA code is stored on each user’s C drive. Application converts text files to pdf and .xls format.   1. RST check against log file data (manual visual comparison) 2. Standard deviation (SD) file check. Manually deleting the lines from the file. DZ: Manual labour, time consuming task 3. L2d file check against log file and against RST, manually and against fault log in log file. (each individual fault mileage is checked manually) 4. Cyclic top file (CYC) check against RST, manually. CYC has start and end location which is checked against RST 5. Dip file check against RST 6. L1d file check against RST 7. Manual task of copying and pasting RST header into all 5 text files listed in 2-6.   Items 2-6 is a manual check of a text file against RST file. Manual comparison of data. At the end of the process there is still no visibility if faults are valid or not.   1. VBA (access) code is used to import text files and convert them to /pdf format. (manual copy of the folder location) 2. Application cross references geography location of the faults against ELR, track ID and mileage 3. Assigns the faults to track engineer and cross reference track information so the right fault is assigned.   Tasks 8-10 will be redundant after new TIGER system is introduced and will not be in the scope of required services. However, information may be useful to bidders.  Process improvement ideas/notes:   * Standard deviation file check is very manual process (deleting lines in text file). Can be fully automated * Text file format is easy to code and automate the processes as long as thresholds are set, and result format is set * SD file comparison with dt1 file can be done in this stage as well in the proposed platform. |

Table 3 Data Integrity Check Stage

#### **Run on run check stage summary**

|  |  |
| --- | --- |
| Stage | Actions |
|  | LR3K conversion tool creates pdf files to be used for reporting.   1. Run on run file comparison with previous historic file. Manual visual check of faults between two files (pdf format) 2. Geo File is used to overlay two traces 3. Comparison of log file with the .l2d file (faults) 4. Invalidation data is written on the piece of paper and entered manually into log file. The text file format is not user friendly for log file entry.   Process improvement ideas/notes:   * Historic and current file can be opened at the same time in the same interface and overlaid for easier comparison * Data values projected on the graph if needed * Automatic recognition in data variation between current and historic file (through raw data file) only highlighting the areas technician needs to focus on * Easy to use functions of; dragging, zooming, panning, toggle data on and off, opening raw data file. * Possibility to attach the function of the graph to log file or any other log system, by clicking on the graph and entering the notes that invalidate or validate the data, hence the location and fault is logged in the log file or system * Aim should be to reduce the unnecessary run on run data checking. |

Table 4 Run on Run Check Stage

#### **CDMS stage summary**

|  |  |
| --- | --- |
| Stage | Actions |
|  | CDMS is a java 5 script application. It is an IT supported software. Microsoft Access code also is used at this stage. Access code graphically compares the dt1 file and standard deviation file (SD).  CDMS require 5 text files and .tpe file.   1. Selection of track sections and marking data as invalid, either for all channels or specific channels. Invalidations made per eighth (220 yards) 2. CDMS creates dt1 file from the SD text file 3. CDMS performs additional invalidation of the SD file against its own created dt1 file 4. VBA Microsoft Access script is used to check the graphic comparison of standard deviation. 5. The data is loaded into the EMS hub for TGR (within 24 hours Small Parse) and TIGER (within 10 minutes of Small Parse)   Process improvement ideas/notes:   * .dt1 file check can be done in one interface and in the same stage as SD file check (CF thoughts?) * Similar process to run on run check, with historic data comparison. Can be integrated in the run on run check (CF thoughts?) as one process. |

Table 5 CDMS Stage

#### **Reporting stage summary**

|  |  |
| --- | --- |
| Stage | Actions |
|  | This stage uses CMD script. CMD script is an IT supported script.  VBA application is also used for log file. VBA application is stored on each user’s C drive.   1. CMD creates invalidation file with a list of faults that technician invalidated 2. VBA reads the files from the run folder and copies pdf reports (from run folder) 3. VBA pastes the data into the server location where track engineers access data 4. VBA creates email. Email needs to be ‘sent’ manually. 5. .tpe.$cdms file created by CDMS is transferred to a folder location for LADS database   Process improvement ideas/notes:   * Email reporting should be left as a final result of the process. However, with the possibility to remove it and send notifications to the software track technicians would be using in the future. Requirement – ability of the app to communicate, exchange data with other software within NR. * Both interfaces used in this stage can be joined within the desired common interface. * Log file is used throughout all stages (except ‘Data Receipt’) of the process. |

Table 6 Reporting Stage

Software and tools used in the process:

* Email
* TRACKSYS converter - converting tpe non-binary data to create pdf trace, csv and geo file, run automatically no input from user (internal NR)
* Log file (excel)
* VBA for extraction of file during cataloging stage
* Zip manual extraction
* PCRST software (internal NR)
* VBA ACCESS LR3K Database script for text file conversion to pdf and .xls (Microsoft access)
* Adobe for pdf run on run graphs check
* Piece of paper for data invalidation checks
* CDMS java script (internal NR)
* VBA script for SD text and dt1 text file comparison (Microsoft access)
* CMD script (Internal NR)
* VBA script for pdf file copy from run folder and email generation
* Email

#### **Software used for data analysis**

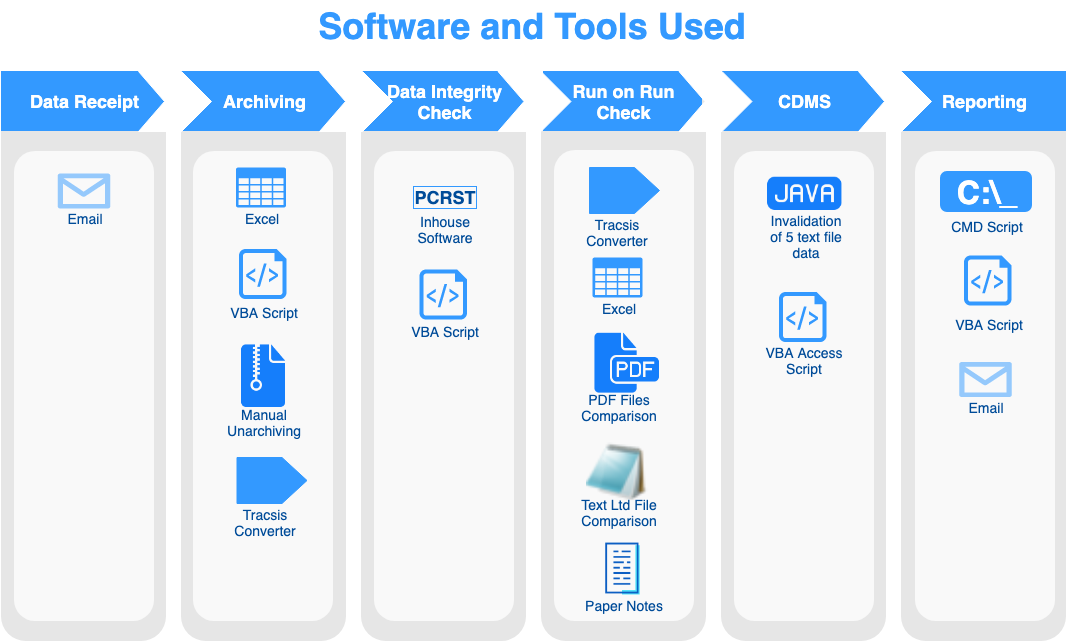


Table 7 Data Analysis Process Software and Tools

Track geometry team main task is to verify and process the track geometry data received from track recording vehicles. Verified data is then uploaded to downstream systems such as CDMS/ TGR and TIGER. However, the data analysis process is manual and time consuming. Process involves multiple software and VBA codes. Also, it involves multiple file formats and conversion from one file format to the other.

Challenges of the current data analysis process

* Data analysis process evolved with stages being added to it instead of being created from the scratch with all processes optimised and customised for a specific purpose.
* No integrated common interface for the data analysis process.
* Data analysis process has too many steps.
* Process uses extensive number of software packages. Sometimes two packages need to be open to compare the data.
* Too many manual tasks that can be easily automated.
* VBA codes still need some manual input that can be automated
* Too many different file formats are used during the data analysis process
* Some data is stored on C drive on each user’s machine with manual updates carried out by each user.
* No consistency in IT support. Some software is supported by IT some is updated within a team (vba codes).
* Multiple folder locations on multiple drives are used during data analysis process. No common folder location.

## **Solution**

NR/AIS are seeking innovative solutions to the challenge described above.

It is expected that a solution to the current challenge will possess the following attributes:

* Automation of manual tasks.
* Ability to read and convert different file formats.
* Recommended solution should be robust and reconfigurable.
* Compatible with the NR IT requirement and operational software (currently Windows 7, to be upgraded to Windows 10 in 2020)
* Scalable to other NR departments with similar activities.
* Easy to use, maintain and develop additional basic coding in house.
* Proposed automated activities should be substantially faster than existing process, allowing staff more time to analyse the data/parameters and classify/report faults.

### Additional information

Reference data sets are available upon request, in accordance with the parameters outlined in the data sharing licence contained within this PIN. This License should be fully signed by an authorised representative and returned to the Catapult, whereupon the data sets will be made available. The document should be completed and retuned without requests for amendment.

An industry day is planned for the week starting the 3rd March 2020, during which further queries and requests for more detailed information can be addressed. If you wish to participate in this event, please provide the names of your representative(s) who will be attending.

The location of the event will be:

Connected Places Catapult  
The Pinnacle  
170 Midsummer Boulevard   
Milton Keynes  
MK9 1BP

**Requirements**

You are requested to submit a response to the challenge statement identified. All responses are required to **be completed within the Word template provided with this PIN and must not exceed   
20 pages (excluding any supporting materials).**

All responses will be collated, and the information will be evaluated in order for us to have a clear understanding of potential development opportunities and the technical readiness of   
possible solutions.