Data analysis for ATO pilot in smartrail 4.0

Vertraulichkeit:	Intern
Zuletzt geändert:	14.10.2019 09:48
Zuletzt geändert durch:	RAO XIAOLU (I-SR40-PMO-PLP)
Dokumenten-Status:	[in Arbeit	in Review	Freigegeben]
Version:	00.01
Dateiname:	ato_3 - data analysis for ato pilot in smartrail 4.0.docx

I.   Änderungsnachweise

Version	Datum	Autor	Änderungshinweise
00.01	2019.09.25	Xiaolu Rao (I-SR40-PMO-PLP)	First draft

II.   Verarbeitete Reviews

Reviewer	Datum	Link Review-Bericht / Verifikationsbericht	Verarbeitung abgeschlossen am/vom
00.01	2019.09.26	\\sbb.sharepoint.com@SSL\DavWWWRoot\teams\p-230\941\Oeffentlich\@Konzeptbericht\Beilagen\Nachweise_Dokumente\ATO_3-Data Analysis\Review Sheet ATO_3 - Data Analysis for ATO Pilot in smartrail 4.0 .xlsx	Jens Nolte (I-SR40-PMO-PLP)

III.   Freigabe

Version	Datum	Rolle	Freigebender / Gez.

Inhalt

1   Motivation

1.1   Data analysis in railway digitalisation

In railway digitalisation, the traditional data analysis and the latest big-data technology has two major applications: rail safety and rail performance.

The safety-information is extracted from the data collected during the monitoring process and the corresponding data analysis. This information can support the decision-making in railway safety management [1] and in accident or incident investigation.

Improve the holistic performance is another major goal in railway digitalisation. Data collection and data analysis are supportive to optimise various railway control systems (e.g. traffic management system, train control system, energy and power management system, etc.), thereby to increase capacity, improve punctuality, improve energy saving, improve customer service and reduce cost.

This document focuses on the potentials of using data analysis to improve railway performance.

1.2   Challenge

For any control system, the feedback data is an essential reference to adjust control decisions. The control decisions can be a new rescheduling plan, a series of speed profiles, different speed recommendations, different traction or braking force, a power distribution plan, etc.

The higher volume and the higher accuracy of the feedback data, the more effective the control decision could be. However, the current quality of feedback data is restricted due to railway fragment and a lack of standardisation[2], which is a major challenge in the progress of railway digitalisation.

Historically, technical and regulatory requirements and operational rules on the railway have been fragmented[2]. Railway fragment brings in two challenges: diversity and isolation.

One of the areas with the greatest diversity is the control, command and signalling (CCS) in railway. For example, although all signalling systems are based on the same block principle, every infrastructure manager adapts this principle to its own operations concepts, resulting in different technical specifications for both mobile and fixed equipment[2]. When the railway systems have such a great diversity, the data collected from such systems has also a great diversity in format and in quality.

Isolation exists in railway products developed by small and fragmented railway suppliers. Every supplier has its own data source and uses its own principle in data collection, data governance and data analysis. Without standardisation, the data share scheme or interfaces cannot be built up between isolated products.

When no qualified feedback data being provided, the suppliers make their own assumptions in their product modelling based on their own experience or theoretical values. Amount of assumptions might lead to unexpected deviations in railway product performance.

However, challenge can also be a chance. The ATO pilots in smartrail 4.0 provides an opportunity to collect data from different railway systems. We try to reduce the diversity and break the isolation by learning the difference data sources and sharing the data analysis results.

In the following paragraph, we summarise the lessons learnt from the ATO pilots in April 2019 and proposes a general data collection and data analysis procedure for the future.

2 Data collection in ATO pilots

2.1 Overview

Data can be collected from different sources. As shown in Figure 1, The data sources can be categorised differently, for example:

- Track-side and train-side,

- IT department, energy department, train department

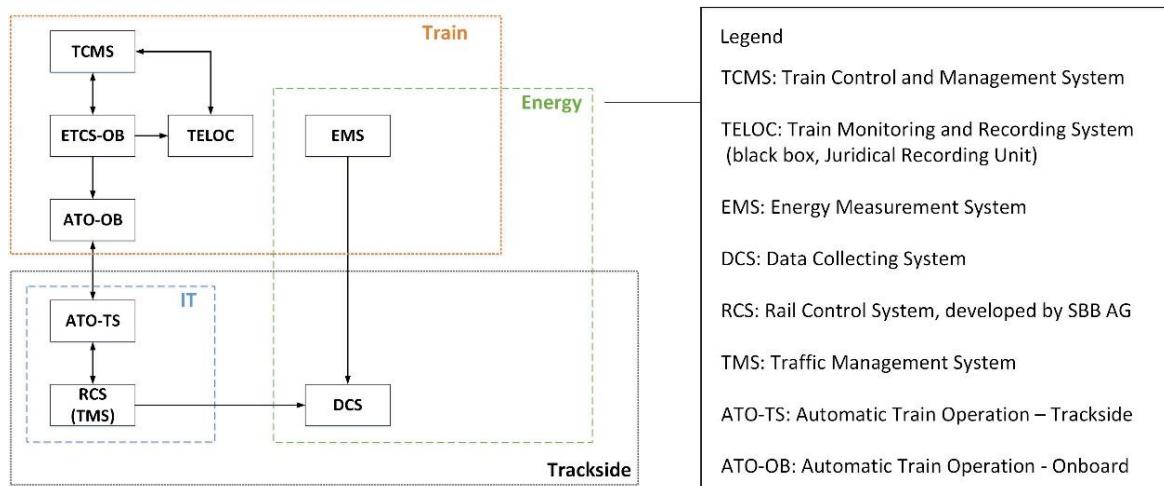- Railway systems: TCMS, TELOC, ETCS-OB, ATO-OB, EMS, ATO-TS, TMS (RCS) and DCS.

Figure 1 The overview of data sources

Data collection procedure can be either automatic or manual, according to the complexity and the design of each railway system.

It is automatic when the data is collected and stored automatically, which means we can collect data after the pilot. It is manual when the data can only be recorded and collected manually during the pilot test.

Based on our experience in the ATO pilot in April 2019, the data collection process is classified as follows:

- Automatic data collection: TELOC, EMS, DCS, ATO-TS and RCS,

- Manual data collection: TCMS,

- Unknown (to be clarified in the future test): ETCS-OB and ATO-OB.

The type of test train in ATO pilot in April 2019 is FLIRT, the Fast Light Intercity and Regional Train, made by Stadler Rail in Switzerland. Figure 2 illustrates an example of FLIRT train, describing the major systems where we can collect data.
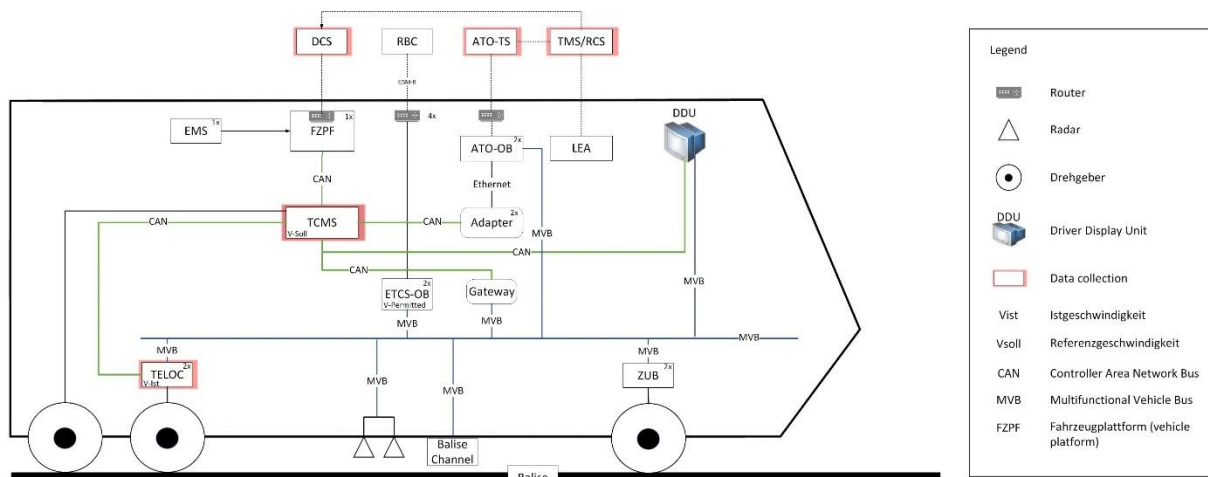
Figure 2 The example of data collection for the FLIRT train

2.2 TCMS

TCMS stands for Train Control and Management System. It is a train-borne distributed system, which comprises computer control units (CCUs), human-machine interfaces (HMIs), Modular Input/output (MIO) devices, Mobile Communication Gateways (MCGs) and the data networks to connect all these together in a secure and fault-resistant manner.

TCMS is often referred to as the "brain of the train" due to its central role in coordinating control and monitoring across disparate systems. The precise TCMS architecture deployed will vary depending on operational requirements and railway market segment, but the purpose and benefits of TCMS are common across all architectures[4].

The main benefits of TCMS is to facilitate technical support to the driver, prompt timely maintenance interventions, and provide an integrated diagnostic and prognostic to minimise or eliminate disruptions.

This report tries to explore the additional benefit of using the TCMS data, which can provide qualified feedback data to support the wayside systems in calibration or optimisation.

In ATO pilot, we collected TCMS data by using the Maintenance computer (UH-PC), which enables us to check the corresponding information recorded in TCMS software. We have interest in the following data: train information, train position, train speed, energy, traction, brake and train control.

As introduced previously, TCMS centralises separate systems. That means, different system might use different parameter to describe the same data information. Table 1 shows all the parameters we tried to record during the ATO pilot in April 2019. This list will be updated for the future pilots.
Based on our experience in April pilot, TCMS data shows a good data resolution of every 100 milliseconds for a record.

Table 1 Example of parameters in TCMS data

Data needed	Example of parameter name in TCMS
Train information	dwFzgNummer
	dwZugNummer
	rFzg1laeng
	rZugsmasse
Train position	dwWegZaehler
	xStandstill
	dwGPSLatitude

	GPSLongitude
	GPSAltitude
	TEL_dTageskm
	rLT_Tageskm
	rDistanz
Train speed	rVistFahrzeug
	rVistAbgleichETCS
	ETCS_wVist
	dwGPSSpeed
	TEL_wVist
	ETCS_wVSollAnzeige
Energy	DP_dwEBezug15kV
	DP_dwEBremsen15kV
	DP_dwEnergieHBU
Traction	rZwischen
	ZB.rZmaxFahrProz
	rZBistZugProzent
	rleistungBezug
Brake	rZwischenBr
	ZB.rZmaxBremsProz
	ETCS_xFzStatusBremsenElektr
	ETCS_xFzStatusBremsenPneum
Train control	IwZBHebelFR
	RIwZBHebelFR

2.3 TELOC

TELOC is a data acquisition and recording system. It is certified to be used as JRU (Juridical Record Unit) for ETCS onboard installations in combination with national data recording.

The TELOC data is downloaded through a software, namely EVA+ developed by Hasler AG, which can provide detailed troubleshooting information for each TELOC, including firmware and configuration version, service logs, etc. EVA+ records data at the same level of details as a juridical recorder, gathering messages during train operation.

TELOC data has different memories for different data sources, such as ETCS data, vehicle data, accident data, etc. Different data sources have different data resolution recorded. During the ATO pilot in April 2019, we mainly checked the TELOC data from the ETCS data memory source where we can collect the permitted speed information and the most restrictive speed information. For the future pilots and their data analysis, we will check more on other data memory sources.

2.4 DCS

DCS (Data Collection System) combines the energy measurement data from EMS (Energy Measurement System) with the train management data from RCS (Rail Control System), as shown in Figure 3. The DCS where we collected data is developed by a company named Microelettrica Scientifica from Italy.

The most important data we collected from DCS includes GPS speed, active consumed power, active generated power and energy consumption.

Currently, the GPS speed from DCS is used as the main reference of actual train states due to its convenient data collection process. However, the data quality of train states needs a further comparison with other data sources, such as TCMS and TELOC system. More discussions can be found in Section 3.1.
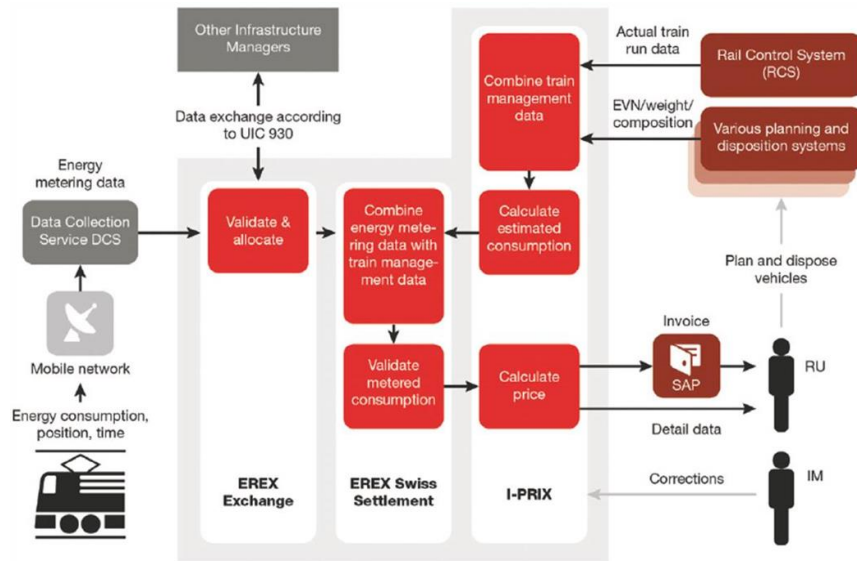
Figure 3 Data flow of energy measurement data and train management data

2.5 RCS

The most important data from RCS is time, especially the estimated departure time and the estimated arrival time at stations and signals.

The train stopping location at stations are not fixed yet, but the signal position is fixed. Therefore, we use the signal passing time in RCS as a reference to compare with the actual train passing time at each signal. The analysis detail is presented in Section 3.2.

2.6 ATO-TS

ATO-TS provides data from Journey Profile (JP) and Segment Profile (SP). The most important data from JP is the traffic schedule at each timing point, such as the latest arrival time, arrival window, departure time, dwell time, etc. SP contains topology information, such as height, curves, gradient, timing point location, etc.

In the future, the analysis of punctuality and energy consumption should take into account the influence from topology variables.

2.7 ATO-OB and ETCS-OB

The data source of ATO-OB and ETCS-OB are not accessible during the ATO Pilot in April 2019. In the future pilots, if these two data sources are available, a more complete overview of data analysis can be achieved.

3 Data analysis in ATO pilots

3.1 Influence of different data sources on train speed profile

3.1.1 Motivation

Train speed profile is an important output delivering from ATO-TS to ATO-OB, representing holistic optimisation objectives, including safety, punctuality, riding comfort and energy efficiency.

Train speed profile is updated according to the dynamic feedback of train states, e.g. train speed and train location. As described in Chapter 2, the train states can be provided by different data sources, such as TCMS, TELOC and DCS, but they are different in data quality, data transfer, data store, etc.

For intelligent railway, the bidirectional communication between the central and the train is a key factor for improving efficiency. Currently the focus is mainly on one direction, that is sending the speed profile from the central to the train. The update of speed profile is mostly based on prediction model, which estimates train states by theoretical train running models. The precision of those models can influence the accuracy of traffic prediction, which further influence the calculation of speed profile as a control reference for the train.

When we start focusing on the other direction as well, i.e. providing relevant data from the train to the central, those train data can be used as a supplementary input to improve the calculation of speed profile or the train prediction model.

In this section, we analyse the actual train speed data in three data sources, DCS, TCMS and TELOC, as marked in Figure 4.
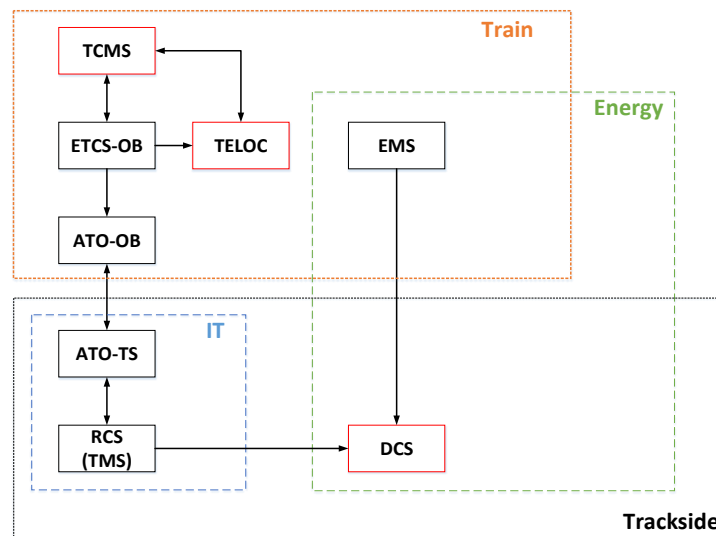

Figure 4 Analyse the influence of different data sources on train speed profile

3.1.2 Speed profile in different data sources

There are three data sources to provide actual train speed data: DCS, TCMS and TELOC. Specifically, DCS records the GPS speed, TCMS records the transferred ETCS speed, and TELOC records the sampled ETCS speed as a juridical record. Besides, TELOC records the most restrictive speed and permitted speed.

The overview of different speed profiles in Location-Speed and in Time-Speed are illustrated in Figure 5 and Figure 6.
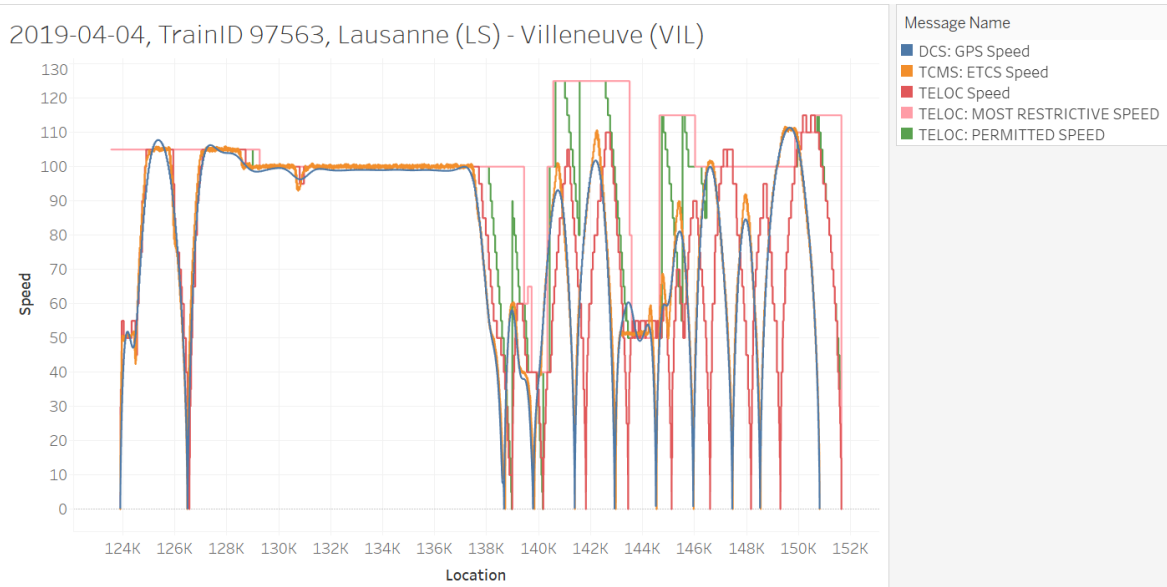
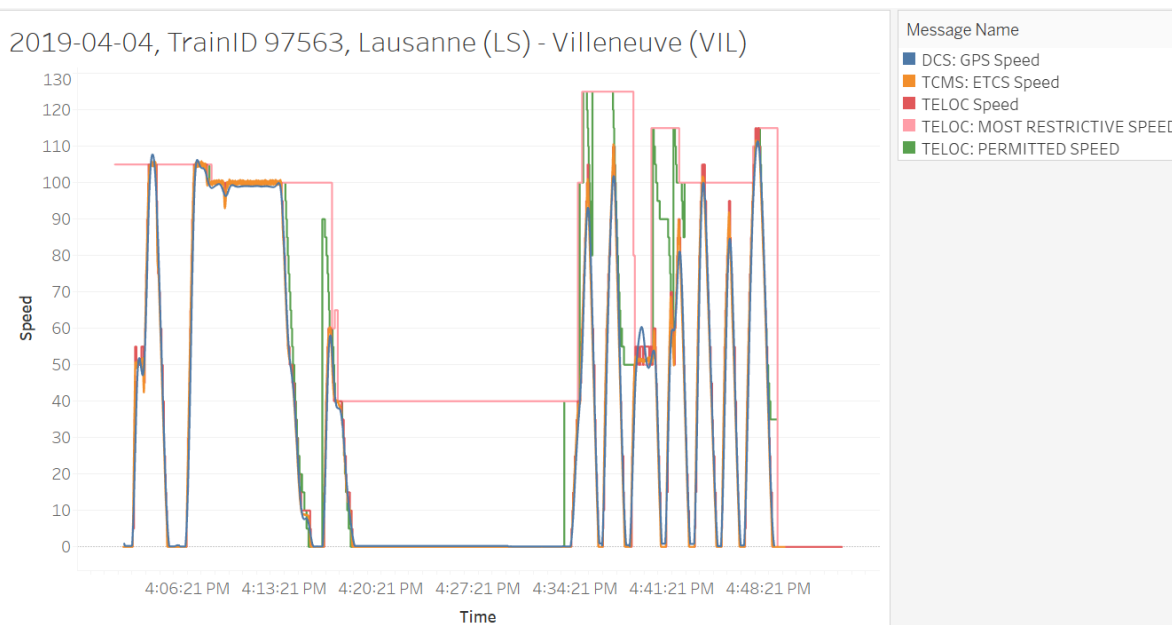*Figure 5 Different sources of speed profile in Location-Speed graph*

Figure 6 Different sources of speed profile in Time-Speed graph

We have several findings in the data comparison. First, the TELOC speed profile shows a location-gap at each stopping location (stations), as shown in Figure 7. The location-gap might be due to an inappropriate adjustment of wheel diameter, but a further analysis shows that the gap is not linear in time or in location. The reason of TELOC data location-gap needs further investigation.

TELOC data has different memory sources. This mainly analysed TELOC data is from the ETCS data memory, which doesn't record the location data, but a rough estimated calcualtion based on speed and time instead.

Later, a further investation found that the TELOC data from the vehicle data memory has location data, which shows a good data quality and data resolution in train speed and train location. Most importantly, the location-gap issue doesn't exist any more, as shown in Figure 8.
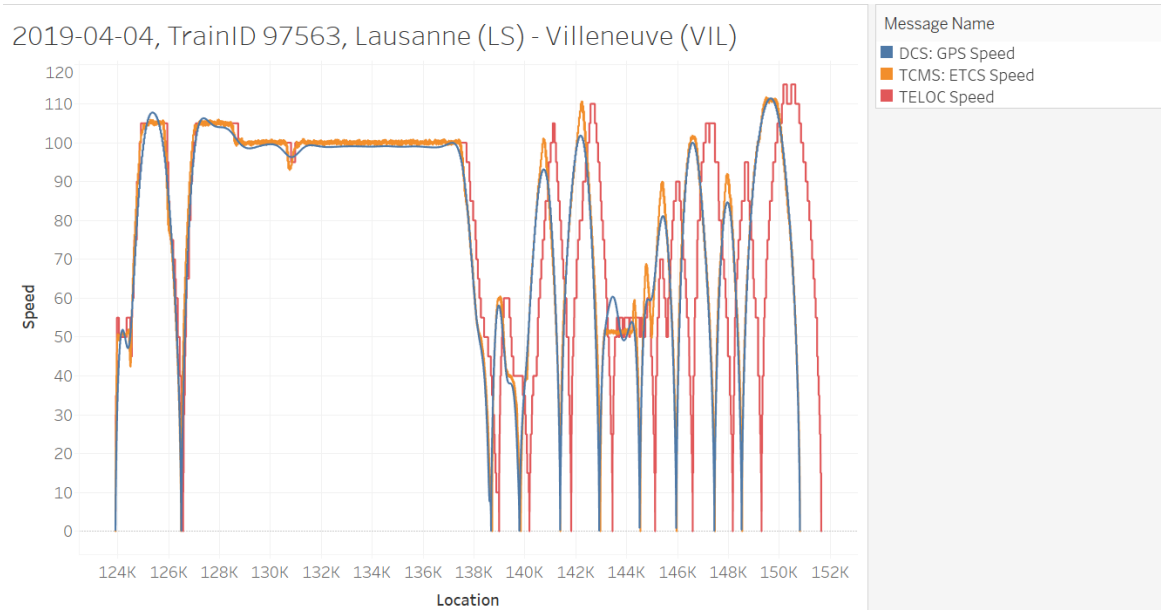
Figure 7 Location gap: Speed profile comparison in DCS, TCMS and TELOC (ETCS data memory) data source
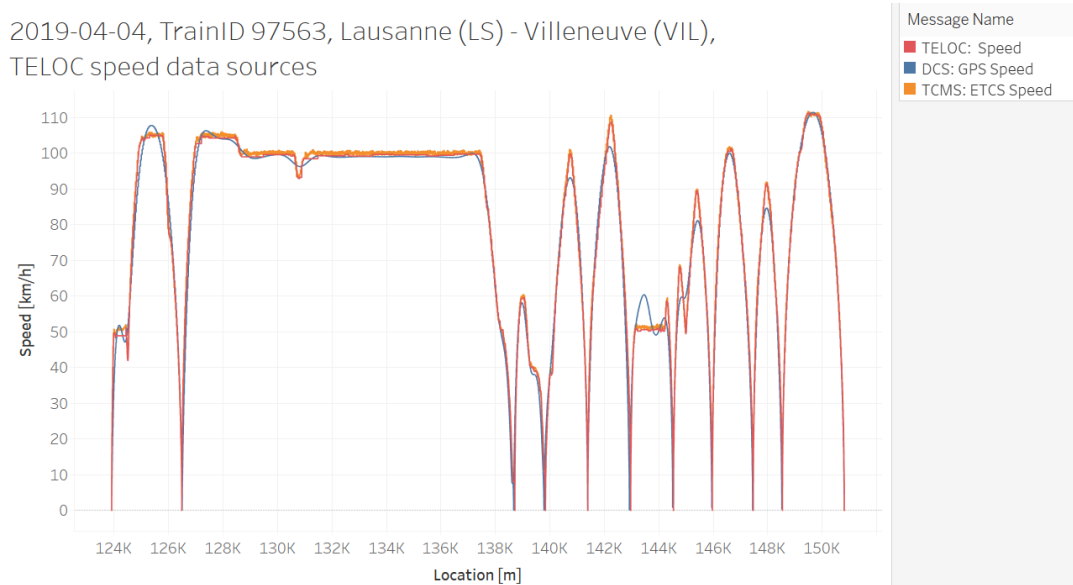

Figure 8 Non-location gap: speed profile comparison in DCS, TCMS and TELOC (Vehicle data memory) data source

Second, the GPS speed (Data source: DCS) and the ETCS speed (Data source: TCMS) show an almost matching location but differs obviously in speed at some locations, as shown in Figure 9.

The GPS speed is filtered by the butterworth filter, while the ETCS speed is a raw data without any filter. The influence of filter using in DCS data needs further investigation.

In addition, there are three short tunnels between Lausanne and Villeneuve, namely Tunnel de Poudex, Tunnel de Bertholod, and Tunnel des Cretes, as shown in Figure 10. The influence of tunnels on GPS speed also needs further investigation.
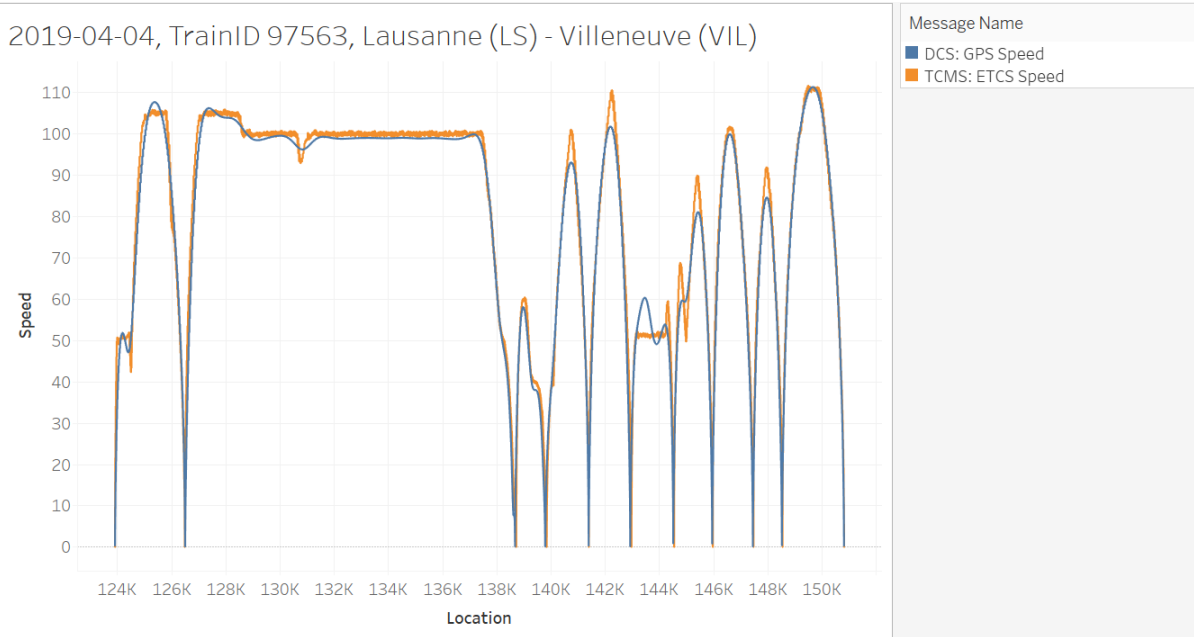
2019-04-04, TrainID 97563, Lausanne (LS) - Villeneuve (VIL)

Figure 9 Speed profile comparison between GPS speed and ETCS Speed

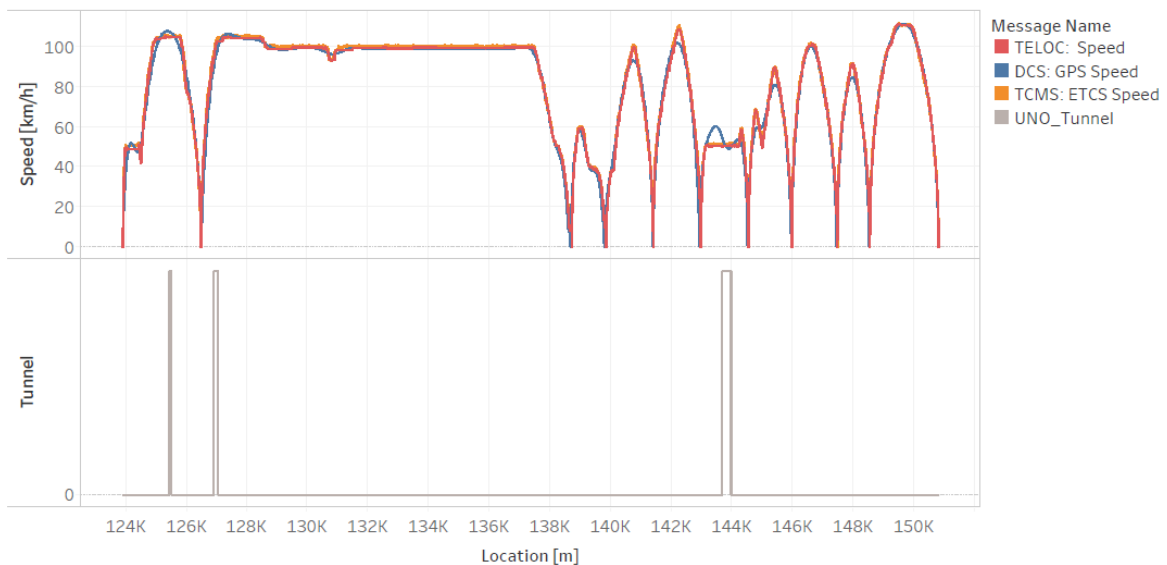2019-04-04, TrainID 97563, Lausanne (LS) - Villeneuve (VIL), TELOC speed data sources

Figure 10 Three short tunnels between Lausanne and Villeneuve

Next, the three data sources show different data accuracy and data resolution. As shown in Table 2. DCS shows a good data resolution in speed and location, while TCMS and TELOC record timestamp detailed into millisecond. TCMS has the largest amount of data record, but which is currently not available for automatic data collection.

One further research could be the analysis of big data technology applying for the historical TCMS data, which the result is not only for predicting the vehicle maintenance but also for optimising the holistic railway operation.

Table 2 Data source comparison in data accuracy and data resolution for one train operation

	Data accuracy (Number of record)	Data resolution (scale of record)		
		Speed	Location	Time
DCS data	1739	10^{-15} km/h	10^{-9} m	HH:MM:SS
TCMS data	28695	10^{-6} km/h	1m	HH:MM:SS.SSS
TELOC data (ETCS data memory)	2062	5 km/h	1m	HH:MM:SS.SSS
TELOC data (Vehicle data memory)	4722	10^{-2} km/h	10^{-1} m	HH:MM:SS.SSS

Furthermore, TELOC data records the speed limitation information, the most restrictive speed and the permitted speed. Analysing the gap between actual train speed and permitted speed in braking could support the analysis of capacity, while analysing the gap between actual train speed and the most restricted speed could support the analysis of safety.

Figure 11 shows all the braking phases during the train running from Lausanne to Villeneuve. If we use the ETCS speed as the actual train speed, this trains follows the restriction of permitted speed and the most restrictive speed mostly. The overspeed part is within 5km/h. Considering the difference in data accuracy and data resolution of two different data sources described in Table 2, this gap in general in acceptable. However, this analysis is only for one train operation, more investigation will be taken in the future pilots.
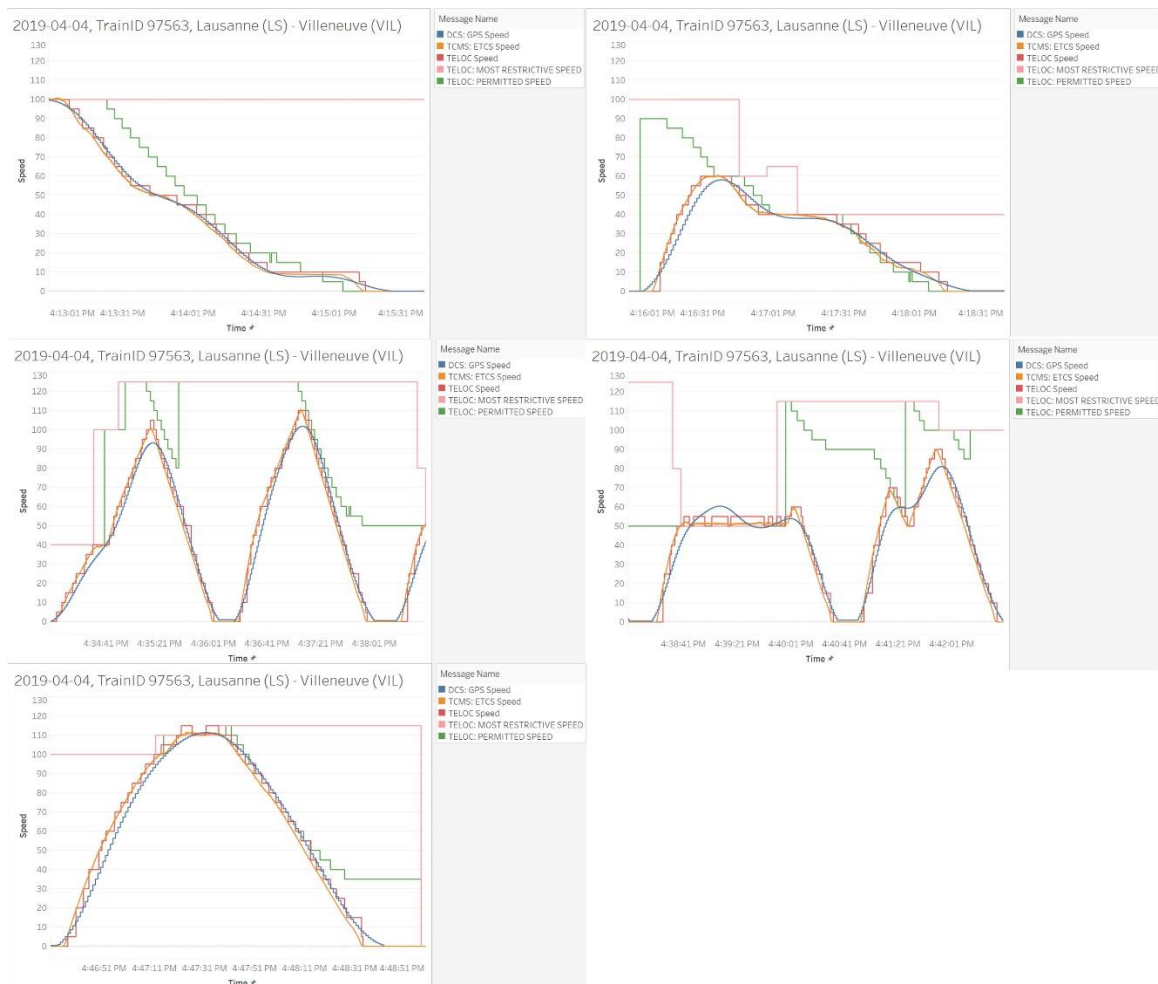

Figure 11 Speed profile in braking

3.2 Accuracy of signal passing time

RCS predicts the train passing time at each signal position. Figure 12 describes the predicted signal passing time from RCS and the actual train passing time from DCS. The gap in between indicates how well the train can execute the RCS plan at signal level. In other words, it indicates how well the ATO-OB can execute the train speed profile from ATO-TS. This result is part of punctuality analysis as well.

Figure 12 illustrates just one example during the pilot, where there was 22 minutes maximum delay. Additional log data is needed for explanation of such a big delay, more comparison will be carried out for the future data analysis, which the result can support to improve either RCS plan or ATO-OB execution.
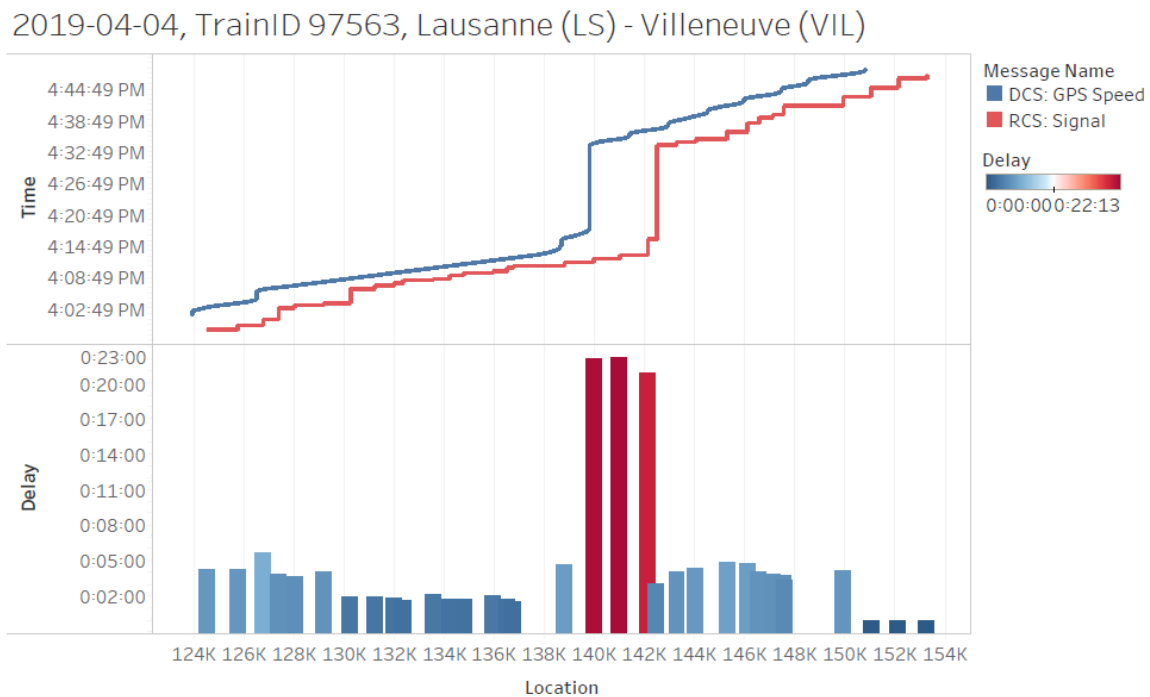

Figure 12 Evaluation of delay at each signal

3.3 Energy and power consumption during train operation

DCS data source provides not only the GPS speed information but also the power and energy related information. As illustrated in Figure 13, the red colour describes the active consumed power during train acceleration, while the green colour describes the active generated power during train braking. Figure 13 provides details of energy consumption during each step of train operation.

In the future analysis, the topology information should be added to further explain the behaviour of train for energy consumption.
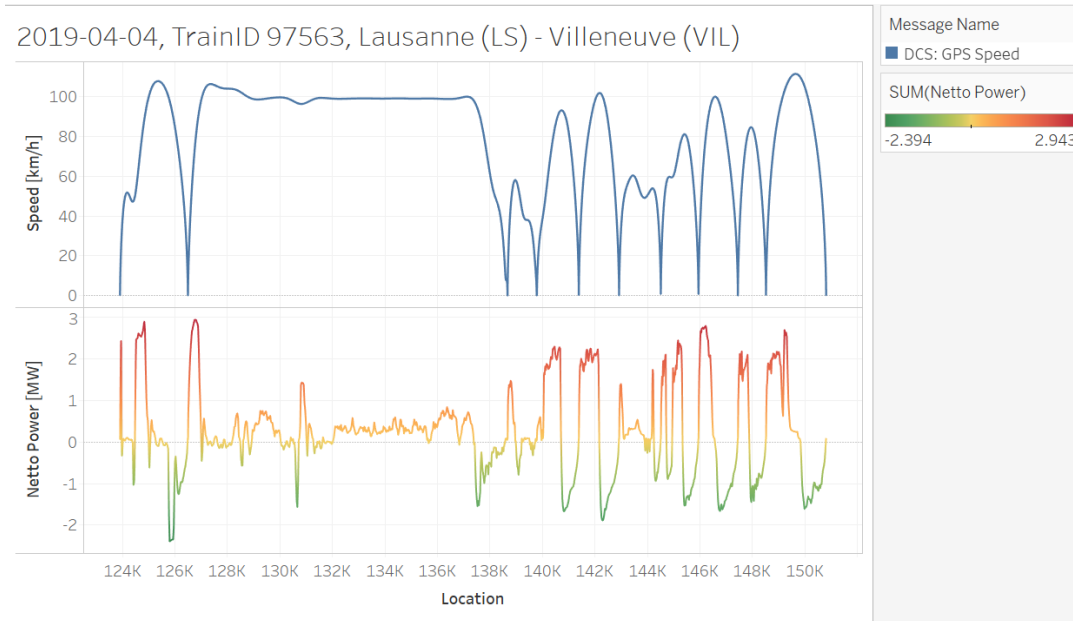
Figure 13 Power consumption during the train operation

4 Future Vision

4.1 Overview of ATO pilots

Table 3 gives an overview of ATO Pilots for SBB, SOB, ZB and DB Cargo. A general data collection and data analysis process is planned to be carried out for the following four pilots.

Table 3 ATO Pilots overview

	SBB	SOB	ZB	DB Cargo
Pilot	ATO Basic (Phase 1-3)	ATO Pilot	Meterspurbahnen	IP5 Demonstrator
GoA Level	GoA2	GoA2	GoA2	GoA2
Train Protection System	ETCS L2 FS	ETCS L1LS	ZBMS	ETCS L2 FS
TMS type	RCS	RCS	ZB-own	RCS
Vehicle type	FLIRT	NPZ FLIRT	Fink	BT TRAXX
Test line	LS – VIL Ca. 28km	Degersheim-Wattwil / Biberbrugg-Arth-Goldau	To be determined	LS – VIL Ca. 28km
Energy measurement available	Yes	To be determined / Yes	Yes	Yes
Planned test start time	2019-10	2019-12	2020	2020-03

Based on the experience of data collection and data analysis during the ATO pilot of SBB in April 2019, we plan to continue learning different data sources during normal train operation. In this regard, two addition data collection and data analysis work were planned at end of July and at the beginning of September, which are based on the same ATO pilot line between Lausanne and Villeneuve for the normal train operation, as shown in Figure 14.

A general data collection and data analysis procedure will be summarised and prepared for the ATO pilots of SBB in October 2019 and the ATO Pilots of SOB at the end of 2019.
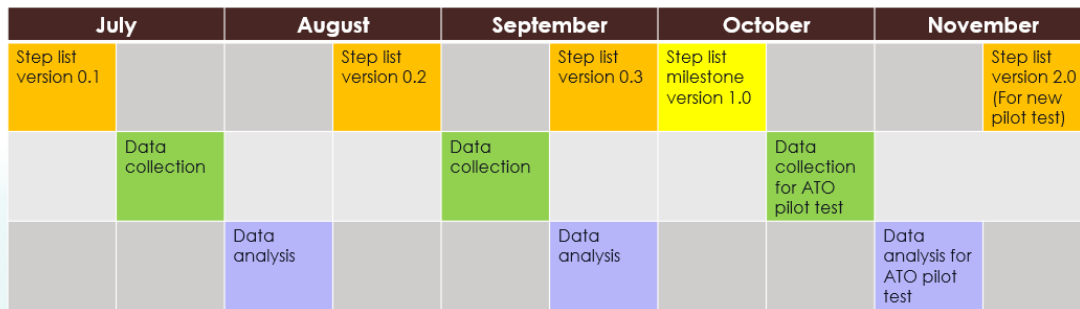
- Step list: A procedure including data collection and data analysis for the pilot test.
- Data collection: Data collection from commercial train operations between Lausanne and Villeneuve.
- Data analysis: A performance evaluation including punctuality, adherence to proposed speed profile, energy efficiency, etc.

Figure 14 Data collection plan based on normal train operation (without ATO)

4.2 Recommended data collection list

To have a general data analysis procedure, we recommend a static data collection list in Table 4 and a dynamic data collection list in Table 5. The discussion of data sources for each pilot project in SBB, SOB, ZB and DB Cargo is in ongoing. Additionally, there would be other useful data as proposed in Table 6.

The usage of those collected data is explained in Table 7, which are used for analysis in punctuality, energy consumption, driving behaviour, etc. These lists will be updated according to experience gained for each pilot.

Table 4 Static data collection list

Data category	Data Index	Data	Data format
Train	1	Train number	Nr
	2	EVN (European Vehicle Number)	Nr
	3	Train length	m
	4	Train mass (mass sum of locomotives and wagons)	kg
	5	vRADN (the maximum speeds for trains of the R, A, D and N series are specified)	km/h (with respect to location)
	6	vMax (the maximum speed of vehicle)	km/h
Trackside	7	Slope (inclination per thousand at the location)	‰ (with respect to location)
	8	Height (topology height at the location)	m (with respect to location)
	9	Signal location	m or km
	10	Tunnel information	m or km
Timetable	11	Planned arrival time (for each timing point)	Time stamp (Date + Time)
	12	Planned departure time (for each timing point)	Time stamp (Date + Time)

Table 5 Dynamic data collection list

Data Index	Data	Data format
13	Time stamp	Timestamp / Date+Time
14	Train position	km
15	Train speed	km/h
16	Energy consumption (mass sum of locomotives and wagons)	kwh (with respect to train position or time stamp)
17	Power	Kw or mw (with respect to train position or time stamp)
18	Train control (including: traction force, electrical braking force, regenerative braking force, mechanical braking force)	N or KN (with respect to train position or time stamp)
19	Acceleration	m/s^2 (x, y, z)
20	ATP log data (e.g. MA, MRSP, Balise, permitted train speed)[ed]	
21	Train Door opening / closing	true or false

Table 6 Other useful data

Data category	Data Index	Data	Data format / Data unit
Others	22	Number of passenegers	Nr
	23	Weather	Nr

Table 7 Proposed analysis based on collected data

Goals of Data Analysis	Data collection in static data	Data collection in dynamic data
Punctuality	1, 3, 9, 10, 11, 12	13, 14, 15
Capacity (e.g. headway time)	1, 3, 5, 6, 9, 10,11, 12	13, 14, 15
Energy consumption	1, 2, 4, 5, 6, 7, 8	13, 14, 15, 16
Power peak	1, 2, 4, 5, 6, 7, 8	13, 14, 15, 17
Driving behavior	1, 3, 4, 5, 6, 7, 8, 9	13, 14, 15, 18
Train Stopping accuracy	1, 3, 9, 10, 11, 12	13, 14, 15
Riding comfort	1, 3, 4, 5, 6, 7, 8, 9	13, 14, 15, 18, 19

5 Conclusion

In railway digitalisation, most discussions of data analysis focus on rail safety, but not enough on improving the holistic performance with the support from data analysis. The ATO pilot projects in smart rail 4.0 provides an opportunity to collect data from different railway systems.

Currently, the quality of feedback data for rail control systems is restricted due to railway fragment and a lack of standardisation, which is also a major challenge in the progress of railway digitalisation. By starting the work of data analysis in smart rail 4.0, we aim to learn how we can use data collection and data analysis to support the rail optimisation at many aspects.

Verzeichnisse

Glossar / Glossar-Referenz

Siehe SR40 Glossar: https://trace.sbb.ch/polarion/#/project/library/workitems/definition

ATO	Automatic Train Operation
ATO-TS	Automatic Train Operation - Trackside
ATO-OB	Automatic Train Operation - Onboard
DCS	Data Collection System
EMS	Energy Measurement System
FLIRT	the Fast Light Intercity and Regional Train
RCS	Rail Control System
TMS	Traffic Management System
TCMS	Train Control and Management System
TELOC	Data recording system onboard

Grafik-Verzeichnis

Tabellenverzeichnis

Quellen / Referenzen

Referenz	
[1]	ERA, "Big data in railways: Common occurrence reporting programme", October 2016.
[2]	Josef Doppelbauer, "Command and Control 4.0", IRSE News July 2018.
[3]	Bernhard Rytz, "Rethinking the heart of railway operations", Railway Gazette International, March 2019
[4]	Simon Tew, "What is TCMS?", Railway Engineer, August 2015.