Operational Principles and KPIs for Self-Organizing Railway Operations

Grant Agreement N°:	875022
Project Acronym:	**SORTEDMOBILITY**
Project Title:	Self-Organized Rail Traffic for the Evolution of Decentralized MOBILITY
Funding scheme:	Horizon 2020 ERA-NET Cofund
Project start:	1 June 2021
Project duration:	3 Years
Work package no.:	WP1
Deliverable no.:	D1.1
Status/date of document:	Final, [2022-June-30]
Due date of document:	[2022-May-31]
Lead partner for this document:	Rémy CHEVRIER, SNCF
Project website:	www.sortedmobility.eu

Dissemination Level		
PU	Public	**X**
RE	Restricted to a group specified by the consortium and funding agencies	
CO	Confidential, only for members of the consortium and funding agencies	

Revision control / involved partners

Following table gives an overview on elaboration and processed changes of the document:

Revision	Date	Name / Company short name	Changes
1	2021-Nov-30	JODEAU J. / SNCF	Initial release
2	2021-Dec-15	CHEVRIER R. / SNCF	Minor revision / intermediate deliverable release with review of indicators
3	2022-June-10	CHEVRIER R. / SNCF	Adding of the operational principles, decision-making objectives and key-performance indicators, final synthesis and appendices
4	2022-June-28	PELLEGRINI P./ univEiffel	Review
5	2022-June-30	BOTTICCHIO N. / univEiffel	Quality Check

Following project partners have been involved in the elaboration of this document:

Partner No.	Company short name	Involved experts
1	univEiffel	P. PELLEGRINI and N. BOTTICCHIO
2	SNCF	R. CHEVRIER and J. JODEAU
3	DTU	
4	BDK	
5	ISTC-CNR	
6	RFI	
7	TUDelft	

Executive Summary

The objective of Deliverable 1.1 is to define the Decision-Making Objectives that will drive the self-organizing traffic and the Key-Performance Indicators which should be monitored and assessed. In addition, this deliverable states the operational principles at the basis of the self-organizing traffic management system.

Chapter 1 is the introduction of the deliverable and reminds the objectives and guidelines of work package 1.

Chapter 2 is a review of the indicators found in the academic literature related to ground transportation (road and railway as well). A classification is then proposed to gather the indicators into specific categories. The indicators which can be computed in real time are also identified.

Chapter 3 presents what operational principles have been retained after interviews with experts from railway infrastructure and traffic management. These operational principles then feed the reflection on the Decision-Making Objectives and the Key-Performance Indicators.

Chapter 4 explains the way of defining the Decision-Making Objectives and the Key-Performance Indicators as well. The reflection relies first on a workshop gathering many railway experts from academic and operational communities (infrastructure managers, railway undertaking, engineering subsidiaries), who were invited to express their ideas, remarks, opinions, warnings. The final choice of the Decision-Making Objectives has been taken by the project partners in meeting.

Chapter 5 gives a synthesis of the work in the deliverable.

Appendix A details the infrastructure managers' answers to the interviews.

Appendix B presents the reflections and thoughts given by the railway experts during the workshop.

Table of contents

Table of abbreviations

DMO	Decision-Making Objective
IM	Infrastructure Manager
KPI	Key-Performance Indicator
RU	Railway Undertaking
TS	Transport System

1 INTRODUCTION

Railway transport systems present many advantages on other transportation modes like, e.g., mass transport, weak energy consumption, safety. In a lot of territories, it represents the backbone of the multimodal transportation networks, whether for passengers or for freight. With an increasing traffic and more and more trains, railway transportation often suffers from many perturbations, which may generate dispatching conflicts, and then delays which propagate and amplify in the whole network.

In urban and inter-urban areas, the current public transport does not fulfil the customers' satisfaction. This dissatisfaction results in a non-ecological and non-efficient increase of car traffic flows in the cities. Therefore, the transport systems must evolve in order to meet the growing desire of citizens' mobility while the urban development must remain climate neutral. Railway transportation may play the role of mobility backbone to accomplish an efficient and demand-aware urban and inter-urban mobility growth.

In parallel, freight transportation also faces operations issues and competition from other modes which tend to make it less utilized for the benefit of long-distance truck transportation in particular. In that case also, railway transportation must evolve to become competitive again and meet the customers' needs. In this way, environmental advantage provided by railways will be enabled.

The SORTEDMOBILITY project aims at building a self-organizing railway operation system. Differently to the historical paradigm which consists in concentrating the whole decisions in an operational center, a self-organizing railway system decentralizes the decision-making concerning regulation to the trains. The main rationale behind this new way of regulating is to make railway traffic more fluid and provide more flexibility and resilience face to disturbances, and finally to transport more users and freight.

Operating self-organizing trains implies a construction phase of the transportation system (TS) relying on three main points enabling the TS to be fair, to exclude the smallest possible share of population and to make its use easy:

1. A high level of service, that means the TS has to be frequent, reliable, demand responsive and resilient,
2. An easy accessibility. To face the increasing number of TS users, its accessibility must become easier for all. The multi-modal use must be simplified and promoted while the TS must spread to reach as many users as possible.

3. A greater efficiency, which is essential to maximize the use of the infrastructure and material. So, it will be possible to welcome more passengers by maximizing the use of the railway capacity. Efficiency is also essential for the competitiveness of multiple actors. The more efficient an actor will be, the more trains will circulate, and the more passengers or freight will be transported.

Operating self-organizing railway traffic is possible thanks to the recent progress in AI, computer science and communication technologies. Indeed, the trains can embed devices allowing them to project and anticipate the future state of the traffic, to share data with the other trains and to take decisions with adapted optimization algorithms. This implies to have fully-fledged autonomous train-agents whose behavior is driven by a utility function which characterizes the maximum benefit for the whole system. That benefit corresponds to the achievement of the three points mentioned before and their evaluation relies on Key Performance Indicators (KPIs) common to the different TS stakeholders such as operators (infrastructure managers, dispatchers, train companies), customers, and urban public authorities.

To be able to define the most suitable KPIs for evaluating, monitoring and driving the self-organizing TS, this document aims at gathering and sorting relevant KPIs for the different stakeholders. Given that their interests are remarkably different, the assessment of the TS will be different following the concerns of the stakeholder under consideration. As a consequence, TS assessment covers a large number of fields such as operations, planning, environment, mobility, economics and others, that we gather into categories.

In addition to those categories, the decision time-horizon is also different with regards to the stakeholders and their practices. The decision time-horizon therefore requires appropriate KPIs. Concretely, operations must be driven by a KPI which will have to be computed in real-time so as to allow the trains to decide what actions to carry out, whereas the daily global performance of the system monitored by the public authorities will be computed at the end of the day on the basis of all operations performed during the day. Customers also have a subjective assessment of some aspects of the TS, in particular the waiting time in the case of the demand responsive transport. This latter kind of indicators may be both monitored continuously but also evaluated after the day of service to adjust some elements of the TS if necessary.

According to the definition of the Sustainable Transportation Indicators, presented in Haghshenas *et al.* (2012), a KPI needs to respect some principles to be effective and relevant. First, it must have a relevant and defined target. Then, it must be measurable, and the necessary data must be available. The indicator must naturally measure what it is supposed to measure and be able to reveal changes and their impact as well. Its transparency is also important: it must be understandable and reproducible for the intended users. Indicators should be as independent as possible from each other. Moreover, in case of comparison, the indicators must be standardized. Jeon *et al.* (2005) gathered KPIs used by sixteen initiatives on sustainable transportation systems and they present all of them sorted in different categories: economic, transportation related, environmental, safety oriented and social-cultural/equity related. Roughly speaking, those categories will be the ones that we will use in our review.

In order to define the Decision-Making Objectives (DMOs) and the KPIs to be monitored and assessed, we followed an approach based on the expertise of railway stakeholders: infrastructure managers, railway undertakings, railway consultants. Thus, beside the review on KPIs, we first interviewed experts from railway infrastructure management in Western Europe to establish the operational principles which will allow the definition of the indicators driving the self-organizing traffic. From there, we organized a workshop with a panel of more than 40 railway academic and industrial experts to express their remarks, opinions, advice and recommendations. Finally, on the basis of all these elements, we chose what indicators will drive the system.

The report is organized as follows. Chapter 2 gives a review of the literature on Key-Performance Indicators covering a large number of fields. The results of the railway infrastructure management experts' interviews and the operational principles are given and detailed in Chapter 3. Chapter 4 introduces and explains the choice of the DMO and the KPIs. Chapter 5 concludes this report with a synthesis on the work done in the Work Package 1. It has to be noted that two appendices complete the report: one about the interviews and the other on the workshop with the railway experts.

2 REVIEW OF THE LITERATURE ON KEY-PERFORMANCE INDICATORS

2.1 Indicators for the performance of the railway system

This part will present the KPIs related to operational performance. Railway management incorporates the real time states and decisions of the transport system. But it has to be noted that KPI related to preoperational steps are relevant to be taken into account.

So, this part will be organized as follows. Firstly, some generic KPIs around the railway operations will be presented. Secondly, KPIs related to the timetable will be discussed. Thirdly, KPIs focused on corrective maintenance will be introduced. Then a focus will be done on KPIs needed for railway traffic operations and for alternative traffic management solutions' evaluation. Finally, an emphasis will be placed on two themes that may have to be integrated for the new TS: Platooning and Decentralization.

2.1.1 Generic indicators on railway operations

The KPIs related to operations can be grouped in different categories. Nevertheless, some are just too generic to be locked in predefined boxes. Here will be presented the most generic indicators.

Vehicles are the backbone of operations. That's why KPIs about their number, their speed or the number of kilometers travelled by each of them appear to be important.

2.1.1.1 Vehicle utilization

The **number of vehicles** in circulation is quoted in almost all scientific articles as a KPI to follow. It is generic and strongly related to the on-field situation. Hansen (2010) mainly discusses about railway timetabling and dynamic traffic management. He almost immediately qualifies the number of vehicles in a period as essential to express the effectiveness of a timetable and the dynamic management of the railway system. Another point of view can be found in Winter *et al.* (2018). The authors' goal is to analyze the performance of an automated demand responsive transport service. Therefore, they study the impact of different measures (vehicle capacity, vehicle dwell time or system costs) on the fleet size. The number of vehicles must be high enough to fulfil the customers' demand at any moment, and especially during peak periods.

The second most important KPIs related to vehicles is their speed. Their maximal speed is always constrained by the network and the equipment of the tracks. So, the **average speed** or the **commercial speed** appears to be more related to the situation on track and are finally different names for the same speed. The definition is quite simple: the speed is equal to the kilometers travelled divided by the travelled time. Nevertheless, the travelled time includes the dwell time at each station. It means that the more the vehicle waits at a station, the lower its average speed. An important consideration on average speed can be found in Lewandowski *et al.* (2017). The authors evaluate a decentralized method to increase ambulances' average speed thanks to smarter traffic lights. The average speed was well identified as the KPI to monitor their progress. The term 'commercial speed' can be found in Georgiadis *et al.* (2014) and in Cascajo *et al.* (2014). Cascajo *et al.* (2014) compare the evolution of different public transport between a reference scenario and a scenario monitored with the measurement. Different KPIs are presented and aggregated to compare the different public transport systems and the commercial speed is one of the first KPIs used to compare them. The authors link the commercial speed to the passenger flow and to operating costs. This link underlines the importance of the average speed. The study from Georgiadis *et al.* (2014) is more generic. The authors design a Performance Measure System to facilitate the evaluation of public transport in Thessaloniki. A lot of KPIs are used and the commercial speed is one of them.

Speed and speed differences between vehicles are important to establish a timetable. Hansen (2010) uses the operating and circulating speed, respectively the speed between any pair of stations and the speed between two consecutive stations, to evaluate the effectiveness of a timetable.

Orth *et al.* (2013) develop some generic KPIs and aggregate them in Level of Service to simplify their visualization. They rely on four main measures: speeds, passenger loads, on-time performance, and headway adherence to build different Level-Of-Service (LOS) assessments. They calibrate their method for the Zurich public transport system, and that calibration enables to precisely define the LOS for the public transport in Zurich.

Then, another important KPI related to vehicles and their utilization is the **number of travelled kilometers per vehicle**. This KPI enables to immediately compare different vehicles and check if some are much more used than others. It can even be used to foresee maintenance operation on the most used vehicles.

Antonialli *et al.* (2019) collect a worldwide benchmark for autonomous shuttle for collective transport. They also provide KPIs to evaluate the operator cost and the quality of service. In this comparison they use the number of travelled kilometers per vehicle.

A slight difference is brought by Sen *et al.* (2011) by considering a number of miles travelled locally. They deliver a huge study on public transit mobility management. They focus on the Texan particularity, nevertheless they provide a full method to evaluate the performance of a public transport management strategy. In this method a lot of indicators can be found to evaluate the efficiency of a public transport. The number of miles travelled locally takes an important place in this study. This number of local miles travelled enables to monitor the accessibility and the livability inside a territory. Better accessibility and livability lead to a better quality of life. This is the reason why Sen *et al.* (2011) use the number of miles travelled locally as a KPI for their study.

The number of vehicles used during a time period is an important KPI. Nevertheless, because vehicles are different, this KPI does not bring any information on the capacity and the maximum flow of passenger that a track or the network can tolerate.

2.1.1.2 Infrastructure utilization

Åhren (2005) mentions the infrastructure **capacity** utilization without describing it. This indicator could be computed as follows: $\frac{ActualCapacity}{MaximumCapacity}$. This indicator allows to see the available margin to increase the exploitation of network capacity by a demand responsive transport system.

Jeon *et al.* (2005) evaluate definitions, indicators, and metrics used to address sustainability in transportation and other infrastructure systems. For this purpose, they collect 16 national studies on sustainability and compare the indicators they used to see their scope. In this collection can be found the **maximum customer flow per hour**. This is related to the maximum number of passengers traveling on the network during a peak period.

Another very important indicator for public transport is the **ticket price** that a customer has to pay to travel on the network. The ticket price is valuable for the user and has a huge impact on competition. Georgiadis *et al.* (2014) use the ticket price to evaluate the accessibility of a public transport system. If the ticket price is too high, customers will not use the TS. We can also quote Eboli *et al.* (2012) where the average cost of a one-way ticket is used to compare different public transports. The ease to purchase such a ticket is also mentioned. It depends on the availability of the automatic ticket machine. However, with a growing digitalization, machines will no longer be needed to buy a ticket. To avoid using the ticket price, Jeon *et al.* (2005) quote the "Affordability of public transit service by lower income residents" used by the Victoria Transport Policy Institute. The fuzzy term appears to be generic enough not to just count the tickets, but all the pass or other techniques used to pay the utilization of public transport.

The number of vehicles, their average speed the number of travelled kilometers, the capacity utilization and the maximum customer flow are more useful for the railway operator. *A contrario* the ticket price has more importance for the users.

2.1.2 Planning and Timetabling

The construction of an efficient timetable is an important step before the operations. The timetable affects the operational performance when not enough or too many trains are planned. Both a lack of trains or too many trains lead to a disturbed situation. One causes a passenger overload, the other creates train congestion. Complete studies on timetables can be found in Goverde (2007) and in Delorme *et al.* (2009). They evaluate the stability or the robustness of a timetable and check the timetable's adaptation to the infrastructure.

The need for a timetable could be discussed. As a matter of fact, a TS based just on real time demand should not plan trains before the demand increases. Therefore, a planned timetable would be useless. Only the demand would drive the train's trip and schedule. Nevertheless, the following indicators can also be used when no timetable is designed. These KPIs remain more useful when a timetable exists.

The first indicator quoted by most articles is the **train frequency** (train per time unit). Train frequency is highly dependent on the infrastructure and the used vehicles. It represents the number of trains passing a point per time unit. The higher the frequency is, the more trains are used, and the more passengers can be transported. References to frequency can be found in Cascajo *et al.* (2014), Djordjevic *et al.* (2016), Mlinaríc *et al.* (2018), Orth *et al.* (2013) and in Proboste *et al.* (2020). Mlinaríc *et al.* (2018) develop several KPIs to evaluate the impact of Intelligent Transport System (ITS) on transport and aggregate them using the Group Analytical Hierarchy Process (GAHP). They say that train frequency is part of "the most significant indicator within the social dimension". The work of Mlinaríc *et al.* (2018) extends the one done in Djordjevic *et al.* (2016). Mlinaríc *et al.* (2018) gather 25 KPIs used to compare the progress made with the installation of an ITS. They also use the GAHP to weight all indicators and report their weight in Mlinaríc *et al.* (2018).Proboste *et al.* (2020) try to define an optimal frequency for Bus Rapid Transit Lines according to a known demand scenario. Then they compare frequency between an open BRT line and a close one. They also express the optimal capacity, the fleet size or the stop spacing as a function of line frequency in the network.

Then, the **regularity** is used to smooth the disparity that can be hidden by frequency. For a same frequency the train distribution during a same amount of time can be very different.

Reddy *et al.* (2009) develop four indicators used by the New York City Transit Authorities to evaluate mass transit. Headway regularity is part of them. In a context of mass transit, the authors highlight that regularity is more important than respecting a predetermined schedule. This indicator enables the penalization of bunched services and gaps in service. Regularity in a time interval is defined as follow. A time interval respects the regularity criteria if every service interval between a leaving train and the scheduled following train exceeds neither 50% of the scheduled interval nor 5 minutes. The track regularity is then expressed as the percentage of trains respecting this criterion.

Schittenhelm *et al.* (2012) describe indicators used by the Danish railways to evaluate timetable. They use the regularity index (RI), computed as follow:

$$RI = \frac{A}{A+B}$$

A = Number of timetabled train paths belonging to a service planned at regular time interval
B = Number of missing train paths that would exist if a service was planned at regular intervals

The regularity index is particularly relevant if different types of rolling stock are used. This definition gives priority to the evaluation of regularity in timetable pattern. They modify this indicator to focus more on the periodicity in a timetable. So, they create the Systematic Timetable Index (STI):

$$STI = \frac{\sum TSP_{mtp}}{TSP_{inv}} * 100\%$$

TSP_{mtp}: Time Span for the most used timetable pattern

TSP_{inv}: Time Span for the investigated time frame

The indicators described in Schittenhelm *et al.* (2012) are more likely to be used during the construction of the timetable than during an operational day.

The last indicator related to timetable is the **dwell time**. The dwell time impacts directly the frequency. The higher the dwell time is, the lower the frequency. Besides dwell time depends on regularity. If a station is not deserved during a long time, more people will get into the train and it will increase dwell time.

In Cascajo *et al.* (2014) the dwell time is an indicator for the service performance. Cascajo *et al.* (2014) describe actions taken in different public transport and their consequences on dwell times. They mainly observe that a reduction of dwell times is strongly correlated with an increase of passenger flow. So, dwell times are an important indicator for the health of a public transport.

Proboste *et al.* (2020) take the opposite point of view and say: "dwell time of each bus at each station is affected by the number of boarding and alighting passengers. Thus, increasing the frequency of a service reduces its dwell times, and therefore increases its operational speed.”.

2.1.3 Real time operations

The management of perturbations is one of the key points of a TS. If there is a perturbation, all must be done to limit its impact on the network, on the planned timetable and finally on the users. This section will gather KPIs related to perturbations and their management.

2.1.3.1 Perturbation evaluation

First, the primary delay must be calculated. The primary delay is the direct delay caused by a technical or environmental disturbance. The secondary delays are the delay caused by a delayed train. Both delays are expressed in minutes or seconds. Because of the difficulty to reduce primary delays, we must focus on the forecast of **secondary delays** according to the actual transportation plan. This KPI enables the comparison on secondary delays between different transportation plans.

A distinction between two types of delays can be found in Delorme *et al.* (2009). The primary delays are induced by some external causes, and the secondary delays are triggered by another delayed train. Delorme *et al.* (2009) introduce stability as the number of secondary delays caused by a primary delay. The proposed method enables the computation of the propagation of secondary delays due to an initial perturbation.

In his Phd thesis D'Ariano (2008) builds different tools to help regulators with the real time dispatching problem. The developed tools aim at minimizing the delay propagation and maximizing the dynamic utilization of the railway capacity. These tools help the regulator to recover as fast as possible after a perturbation.

To compute the secondary delays, their propagation in the network must be analyzed. Goverde (2007) develops a method based on max-plus algebra to compute the recovery time for a define timetable and to compute other statistic on delay propagation. If this work aims at qualifying the adaptation between a timetable and the infrastructure, it can also be used to compute the secondary delays. The secondary delays can be computed either for the actual transportation plan or for a considered solution.

The other indicator linked to delay is the **late arrival of each train** to its terminus. This indicator is here to balance the previous one. It is natural to minimize the overall secondary delays, but this minimization must not be done at the expense of one train. A train must not stop all day long to reduce the overall secondary delays in the TS.

2.1.3.2 Rerouting, Reordering and Rescheduling Trains

The following indicators are more useful to evaluate solutions and to see their theoretical impact on the TS.

The **Number of Relative Reordering (NRR)** is introduced by Quaglietta, *et al.* (2016) to count the number of changes between two different transportation plans. The authors edit new transportation plans each 30 seconds according to the on-track situation. Then, they evaluate how far the new transportation plan are from the actual one. The NRR counts the trains which are not at the same place between the two transportation plans. Therefore, the NRR enables to see how different the transportation plans are. This KPI could be strengthened by using the Disabled Route Ratio described in Wu *et al.* (2018). The Disabled Route Ratio is based on the network's infrastructure and assesses the influence of failed node or link on traffic performance. If the NRR shows differences between different transportation plans, it is not precise enough. So, Sidi *et al.* (2018) present a multi-criteria approach to monitor a transport system. To avoid or correct encountered difficulties, they evaluate solutions with overtaking, short-cut or U-turn. Solutions can slightly differ from the original one. In order to evaluate how different they are and mainly where they are different, the **number of skipped stops**, **number of changes** or **number of impacted vehicles** are counted.

Finally, **cost adjustment** due to a perturbation must be evaluated. Wu *et al.* (2018) define the cost adjustment as the route cost for an OD (corresponding to one among several possible route costs chosen by the user) after the perturbation divided by the route's cost in a normal situation. The cost adjustment allows the identification of where a perturbation will cost the most.

2.1.4 Evaluation of decision paradigm: centralisation vs decentralisation

Currently, most transport systems are organized in a centralized way. It means that almost all decisions are made in a single place that gathers all network information. Nevertheless, a decentralization approach is now emerging. This decentralization will allow decisions based on local information. Marcelli *et al.* (2020) wrote a state of the art on decentralized approaches for railway traffic management in order to reduce delay propagation. Decentralization in railway transport is often linked to multi-agent system (MAS). Indeed, all vehicles and agents in the system can bring information and make decisions. The multiagent system research field cares about the relationship between the agents.

Lewandowski *et al.* (2017) use a decentralized approach to ease the circulation of ambulances by dealing with traffic lights. The traffic light enables or forbids the circulation according to the waiting cars and ambulances number, and the neighbors situation. This study shows that ambulances travel time decreases and that their average speed increases. Lewandowski *et al.* (2017) also observe that the situation does not deteriorate for other vehicles.

Perrachon *et al.* (2020) describe the feasibility of solving railway conflicts in a decentralized way. To compare centralized and decentralized strategies. They use quality of service defined as the ratio between the total passengers' travel time computed after simulation and the theoretically possible one. The total passenger travel time remains similar between centralized and decentralized operations. So, they conclude that decentralization is as least as good as centralization to solve railway conflict.

Van Heeswijk *et al.* (2018) compare a centralized and a decentralized method for the timetable problem applied to road freight transportation. Their ADP (Approximate Dynamic Algorithm) performs very well on small instances and outperforms the centralized policies on larger instances. They study the dispatching problem with a centralized and a decentralized learning method. The centralized policy does not find solution for instances with 10 and 25 agents, while the decentralized one finds solution for these instances and seems to have the possibility to scale with larger instances. The metrics used for comparison are the weighted reward, the traveled transport distance, the delivery time and the number of successful deliveries.

The **decision-making time** plays an important role. Ren *et al.* (2007) describe different consensus strategies to set a common rendezvous point between vehicles. They assume that the vehicles' information state does not evolve throughout time. To ease this hypothesis a quick decision is necessary. If too much time is needed, the situation will be way different, and the decision may not be suitable to the new situation. Therefore, the decision-making process must be quick enough. However, the conflict must be solved, and the system should not spend too much time solving it. If the decision-making is quick, but the made decision is bad and solves the conflict too slowly, other problems could be faced.

The decentralization needs to select a decision among the proposed ones. Two ways are listed as possible research directions in Marcelli *et al.* (2020): auction and "coopetition".

Stojadinovìc *et al.* (2019) design a hybrid auction system to assign train paths in highly dense areas. The proposed method enables the maximization of the income of the infrastructure manager. Vernazza *et al.* (1990) define a priority value to sort trains according to their importance.

The priority value defined in Vernessa *et al.* (1990) is used in a MAS to help local decision centers to organize train transit according to trains' characteristics. The priority value can be considered as an indicator which gathers the importance of a train, its schedule's adherence, or its induced delay propagation. Then, the train with highest priority value goes first.

Both designs have in common that a higher auction (or priority value) is the sign of a very demanded train path. So, the higher the values are, the more conflicted the situation can be. Thus, the **auction value** and the **priority value** must be monitored.

2.1.5 Evaluation of Platooning use on the whole performance

During the nineteen-forties, platooning was first considered by car's manufacturer. Kavathekar *et al.* (2011) define platooning as the creation of "platoons or linked vehicles which will travel along the Automatic Highway System acting as one unit". Thanks to this technique, space can be saved on highway by reducing the inter distance between vehicles. All vehicles follow the leader, their passenger can run their lives as they want and save fuel. The survey in Kavathekar *et al.* (2011) gives an overview of the literature on platooning published between 1994 and 2010. Platooning seems to be interesting because it enables to reduce the space between vehicles, to increase the capacity of a track and to save energy. In the case of railway, platooning will just concern vehicles which share the same portion of rail during a trip. To evaluate the influence or the gain of platooning in the new transport system some KPIs need to be introduced.

One of the first Key Performance Indicators to be considered is the **convoy's speed**. A convoy must adapt its speed to its slowest member; therefore, vehicles cannot always drive at the highest speed. The study in Khan *et al.* (2005) grants speed a lot of consideration and provides a decision-support algorithm to adopt the most efficient speed for all vehicles. When a vehicle decides to leave or join a convoy, its speed must evolve to follow its decision. Thus, the speed of convoys is a major concern for the platooning's technique.

As said previously, a convoy must adapt its speed to the slowest vehicle. Therefore, the **number of vehicles in a convoy** is an important indicator. Traveling in convoy presents more benefits than traveling alone. So, as many vehicles as possible should be part of a platoon. Nevertheless, because safety remains a key point of traveling, the number of vehicles in a convoy must depend on the communication performances of each vehicle. Karoui *et al.* (2017) discuss about a nominal and a degraded mode of communication according to the number of vehicles, the precision of their information and their communication. So, the **size of each convoy** must also be monitored, as well as the number of vehicles in a convoy.

Platooning is better than lonely trip, but the benefits must be computed to be sure that platooning is worth it. Kavathekar *et al.* (2011) mention a reduction of 55% of the drag coefficient for the second to fourth vehicle in a convoy, leading to a reduced consumption. Khan *et al.* (2005) discuss a lot the benefits that a vehicle has by entering a convoy and that a group of vehicles must have if they split the convoy to drive faster. Yasin *et al.* (2020) work on the energy save by vehicles' sensors when relying on a leader in a swarm. Only the leader turns on its sensor's, he warns the other vehicles if an obstacle arises. The followers will turn on their sensor just to avoid the obstacle before turning them off and relying on the leader. Therefore, the **energy saved** must be controlled as the main benefit brought by a convoy.

Platooning seems to be a promising way to save vehicles' energy by relying on a leader-follower relationship. Nevertheless, all gathered and found papers are focused on platooning for road vehicles. Platooning for rail vehicle seems not to be a concern for the current research. However, the number of vehicles in a convoy, their speed or the saved energy seems to be the most important KPIs in order to evaluate the influence of platooning in a transport system.

2.1.6 Measuring the impact of corrective maintenance

The subject of maintenance is on the edge of railway production. It must be mentioned that predictive maintenance does not belong to our scope. Its schedules are planned in advance, and the concerned tracks or materials are not available for the operation during the maintenance performance. All KPIs presented further concern corrective maintenance.

Stenström *et al.* (2012) provide indicators to help maintenance organizers to see the consequences and the benefits of a maintenance work. The licentiate thesis of Åhren (2005) gathers 25 KPIs used by the national Swedish railway and by the Banverket Northern Track Region. This part will focus on the KPIs that these articles have in common.

The most important indicator is the **cost of a maintenance work**. The cost can be divided in different parts as the contractor cost, the energy needed for the repair or the use of a specific material. Stenström *et al.* (2012) even mention the cost for indirect maintenance personnel. Åhren (2005) decides to evaluate the maintenance cost per track kilometer. Nevertheless, the kilometer's number is directly related to the network's size. Thus, this KPI is standardized and is more useful for a comparison between different TS.

The other very important indicator for the maintenance is the **repair time**. Åhren (2005) considers just the time needed for the repair. While Stenström *et al.* (2012) also count the waiting time for the repair to begin. Indeed, the time for the workers to go to problem location can be significant.

So, two indicators concerning the maintenance seem particularly interesting: the cost of the maintenance work and the duration between the identification of the incident and its repair.

2.2 Mobility

In this section the KPIs related to mobility will be focused on. Gabriel Dupuy, in Dupuy (2011), characterizes mobility with three criteria: immediacy, instantaneousness and ubiquity. Immediacy is the TS's capacity to be available without delay after its solicitation. Instantaneousness is directly related to the quickness of the transportation; it concerns the duration between the moment when the customer is picked-up and the moment when he is dropped off. Ubiquity is only a spatial concern. It characterizes the access to the TS wherever we are. Therefore, ubiquity is strongly related to the network's geometry and has, with this definition, no impact on operations.

In addition to these main criteria, reliability, which is not fully integrated in the immediacy's definition, seems necessary to evaluate the operational performance of a TS. The comfort is also added to take into consideration the quality of on-time transportation.

Thus, this section is divided into four parts: instantaneousness, immediacy, reliability and comfort.

2.2.1 Instantaneousness

As explained previously, instantaneousness is linked to the transportation duration. Nevertheless, several indicators can be used to qualify and bring precision on instantaneousness.

The first indicator is the **travel time (TT)**, which is defined as the duration between the departure and the arrival.

The TT is used by almost all articles on public transport and remains one of the main concerns of transportation's policies. Here are articles using the TT: Djordjevic *et al.* (2016), Karlaftis *et al.* (2012), Nesheli *et al.* (2017), Nesheli *et al.* (2015), Cao *et al.* (2014), Payne *et al.* (2015), Mlinaríc *et al.* (2018), Antonialli *et al.* (2019), (Sen *et al.* (2011) and Krmac *et al.* (2017).

Djordjevic *et al.* (2016) and Payne *et al.* (2015) evaluate the impact of Intelligent Transport System (ITS) on transport. In order to compute the brought improvements, they use the Travel Time.

Mlinaríc *et al.* (2018), Antonialli *et al.* (2019) and Sen *et al.* (2011) are huge studies on public transport, autonomous shuttles and on Texan mobility. All their authors use the TT to evaluate the TS situation and to compare it with other TS.

Corman *et al.* (2017) qualify the travel time by considering only the time spend in public transport. It means that waiting time or time spent in connections is not taken into consideration. The **in-vehicle Travel Time** introduced in Basu *et al.* (2018) indicates quite the same concept. Basu *et al.* (2018) decompose the total travel time into in-vehicle travel time, waiting time and walking time.

The **Average Additional Travel Time** is defined by Van Oort (2011). This indicator enables the expression of service variability throughout the distribution of passengers' travel time. The AATT is computed as follow. The scheduled TT is deducted from the average travel time to get the AATT.

$$AATT = \overline{TT} - T_{scheduled}$$

The average additional travel time is not commonly used. The survey on service reliability conducted by Van Oort (2011) shows that only London (out 23 answers) uses a comparable indicator: the Excess Journey Time. The Excess Journey Time compares actual and free-flow travel time instead of actual and scheduled.

The **running ratio** is introduced by Kaparias *et al.* (2011) to compare cars and public transport. The running ratio is defined as the total travel time divided by the time spent in an in-service vehicle. This ratio underlines the useful time spent in a vehicle. If the vehicle is stopped, or if the customer is waiting for a connection, the ratio goes up. The higher the ratio is, the more unefficiently time is spent by customers.

The **dwell time** used in Cascajo *et al.* (2014) represents the time spend in a stopped vehicle. This dwell time in public transport is mainly spent to enable boarding and alighting of passengers. Therefore, this dwell time represents the time's difference between a shared vehicle and a personal one.

Wang *et al.* (2010) introduce the "mobility", defined as the travel time divided by the Euclidean distance between origin and destination (OD) for all possible ODs. They apply this indicator to a case study in Mississippi. They study not only mobility but also reliability, security and environmental impacts. The use of geographic distance is justified because companies does not care about the route traveled by vehicles. Their goal is to transport passengers from their origin to their destination regardless of the travel path. Nevertheless, because their "mobility" is in $s.m^{-1}$, it may be better to consider its inverse to have the same dimension as a speed.

2.2.2 Immediacy

The immediacy is the quality of bringing someone into direct and instant involvement with something. In our case immediacy is strongly linked to the duration between the moment someone wants to be transported and the beginning of his/her transportation.

The first immediacy's KPI is the **waiting time**. The waiting time is described in Antonialli *et al.* (2019), Basu *et al.* (2018), Trompet *et al.* (2011), Reddy *et al.* (2009). Basu *et al.* (2018) compare Autonomous Mobility on Demand (AMoD) and mass transit system in different scenarios. The scenarios are designed to evaluate the importance of each transports' mode. They observe that the AMoD has excellent results but that the mass transit system remains crucial to avoid congestion during the morning and the evening peak. The waiting time is one of the two user-centered indicators used to evaluate the utility of each compared mode.

Antonialli *et al.* (2019) compare the quality of service between different public transport systems using autonomous shuttles. Waiting time is one of the five selected user-centered KPIs. They choose 9 key performance indicators to cover all typologies of use. However, they plan further studies to evaluate the usability of their KPIs and to better understand if more KPIs are needed to evaluate the business model of autonomous shuttles.

Trompet *et al.* (2011) monitor the Excess Waiting Time, showing that waiting time is an importance concern for immediacy. Nevertheless, the Excess waiting Time seems more likely to be linked to reliability.

Su *et al.* (2019) develop an agent-based model in order to evaluate the boarding and alighting efficiency of autonomous public transport. They aim at computing easily the boarding and alighting time without building a rail vehicle to run test on. Their method enables to simulate the boarding and alighting of passenger according to the geometry of the vehicle the behavior of passenger, the level of crowding or the employed strategies of waiting customers. The boarding time can then be simulated with a deviation of less than 11% comparing to real-world experiments. They use three metrics: the alighting time, the boarding time and the settle time (time when all boarding passengers have found a place).

Winter *et al.* (2018) use rather the idle time than just the waiting time. The idle time is the waiting time plus the time spent in the vehicle before its start. The passenger idle time is presented as the user convenience's indicator. They use it to build an efficient Automated Demand Responsive Transport Service. Passengers wish to travel through a public transport network and are conveyed by small size vehicle during the morning peak hour. Winter *et al.* (2018) aim at minimizing the necessary fleet size while minimizing the operational costs. The optimization does not run on user convenience indicators, but they are still monitored.

Nesheli *et al.* (2015) focus their consideration on waiting time during connections. They use 4 predefined tactics (hold and skip, hold, short turning and none) to reduce the transfer waiting time for the user. Their techniques modify the train's departure time or arrival time leading to a more synchronized timetable, even in case of perturbations. These techniques used in real time show a slightly impact on global travel time, but the number of direct transfers (without waiting time) increases drastically.

2.2.3 Reliability

The reliability is a mainstay of customers' trust and a key point for the offered quality of service. Reliability has different aspects: punctuality, which indicates the respect of a scheduled plan, and the variability, which characterizes the travel times differences between same OD's travel. Van Oort *et al.* (2009) show that customers perceive the train frequency differently according to the irregularity, whether it is closer or farer. The perceived frequency decreases with the irregularity and with it the customer's satisfaction. Therefore, reliability must be monitored. Van Oort (2009) conducts a survey on public transport reliability, discussing how can it be quantified, and which design could affect it. Almost 30 authorities and operators in 15 countries answer the questionnaire. 74% of the cities use a bandwidth to monitor schedule adherence, while 21% use the average punctuality. Van Oort (2014) presents punctuality and regularity KPIs. He underlines that the importance of these indicators increases with the demand at a station.

2.2.3.1 Punctuality

Punctuality is an important part of reliability, but punctuality can only be evaluated in TSs with a timetable. As delays appear when the actual schedule differs from the planned one, delay and punctuality are strongly correlated.

Famurewa *et al.* (2014) measure punctuality with either minutes of delay or the number of trains that arrive earlier or later than scheduled. They also precise that the philosophy of punctuality can differ from one regulator to another. Cascajo *et al.* (2014) expresses punctuality as a percentage of on-time buses, which implies the need to compute the number of delayed vehicles.

Cao *et al.* (2014) and Trompet *et al.* (2011) use the qualification "standard deviation" to compute and total absolute delays between all ODs. The use of absolute delays implies that if a train arrives before its scheduled arrival time, it will contribute positively to the absolute delay, as well as if it is late. The same indicator is used by Van Oort (2014), he calls it the average deviation from timetable at a specific stop, or a set of stops.

The **sum of all delays** is also used by Djordjevic *et al.* (2016) to monitor the efficiency of ITS. They link the efficiency of TSs to delays. Eden et al. highlight that the increase of public transports reliability and the reduction of delays can work as a ”pull-factor” and encourage the arrival of new customer in public transport. A punctual transport system appears more attractive for customers.

Lüthi (2009) evaluates in his PhD thesis the impacts of a real time rescheduling framework on the knock-on delays. The knock-on delays and their propagation through the network were significantly reduced with full functionality of their integrated real-time rescheduling framework. Bouvet-Agnelli (2016) uses the average punctuality to evaluate the Transilien L line. According to Van Oort (2011), 5 operators out the 23 interviewed use this indicator. The average punctuality (AP) is expressed as follow, which is a user punctuality and not a train one:

$$AP = mean_{users}(UserArrival_{actual} - UserArrival_{scheduled})$$

The AP depends directly on the users' number, the more they are, the highest the AP could be. Bouvet-Agnelli (2016) presents also the users' punctuality as the percentage of users which are less than 5 minutes late. This indicator takes into consideration that the more crowded a train is, the more users its delay will impact. The users' punctuality is also used by *Île de France Mobilité* to monitor commuter trains managed by *SNCF* and *RATP*.

Trompet *et al.* (2011) use delay associated to a threshold. They count the number of vehicles whose delay is superior to 5 minutes. Then they express this number as a percentage. This method has the merit to limit the impact of inaccuracy in time measure on delays. However, the threshold must be defined according to the regulator policy.

Adeline *et al.* (2017) define punctuality as the “ratio of train trips delayed by less than *x* minutes”, their formula follows the standardized definition of the UITP. They use this indicator to evaluate their simulator based on stochastic time Petri Nets. It enables to simulate quickly the behavior of a metro line. Nevertheless, the simulation does not take into consideration the passenger flow and draws randomly the dwell time and the trip duration according to predefined distributions, which limits the realness of a situation. Van Oort (2014) describes a more generic indicator, the ratio is computed not only by train delayed by less than *x* minutes, but by train which arrived neither earlier than *y* minutes nor delayed by *x* minute.

2.2.3.2 Variability

The travel time variability is also a way to monitor the reliability in public transport.

Kieu *et al.* (2015) study the Public Transport Travel Time Variability for buses. They define KPIs around it to monitor different variations. The travel time can be different depending on the vehicles used or the time period they travelled. During peak period travel time could be smaller than during off-peak period.

Bouvet-Agnelli (2016) uses the range variation, defined by Tu *et al.* (2007), this indicator shows how steady a travel time between an OD can be:

$$LV = F_{TT}^{-1}(90) - F_{TT}^{-1}(10)$$

$F_{TT}^{-1}(x)$ designates the x^{ith} percentile of the travel time for the observed OD.

Other range variation can be computed between the 80^{ith} and the 20^{ith} percentile or the 70^{ith} and the 30^{ith}, all are reported in Bouvet-Agnelli (2016). Nevertheless, this range variation cannot be computed in real time because the computation of percentiles is needed.

Trompet *et al.* (2011) use the Excess Waiting Time, with three other KPIs, to compare different bus companies on service regularity. The excess waiting time is defined as the difference between the Actual Waiting Time and the Scheduled Waiting Time.

$$EWT = AWT - SWT$$
$$AWT = \frac{\sum_{i=1}^{N} AHway_i^2}{2 \times \sum_{i=1}^{N} AHway_i^2}$$
$$SWT = \frac{\sum_{i=1}^{N} SHway_i^2}{2 \times \sum_{i=1}^{N} SHway_i^2}$$

Where AHway means actual headway and SHway scheduled headway. The excess waiting time represents the non-scheduled waiting time for a user, which is directly linked to a variability's increase.

Adeline *et al.* (2017) introduce also the regularity based on the respect of headways. Regularity is the "ratio of train departures at a specified station complying with planned headways within *x* minutes, over the total number of departures at that station". Their formulation allows choosing which stations have to be monitored to focus only on the most important ones.

Van Oort (2014) introduces the coefficient of variation of actual headway of a line *l* at a stop j.

$$Cov(H_{j,l}) = \frac{StD(H_{j,l})}{E(H_{j,l})}$$

The StD represents the standard deviation, $E(H_{j,l})$ is the expected headway from line *l* at stop j. This indicator neglects the number of passengers and focuses on trains. The respect of a timetable is not important, only counts the headway's regularity. The coefficient is equal to 0 if all trains are separated by the same headway.

2.2.4 Comfort

Also, comfort is a concern for users. If traveling is exhausting, users will less likely use public transport. Comfort can be linked to soft seating, nevertheless, to consider soft seating the users need to be seated. More general indicators are the crowding in stations and cars.

The **user number** must be monitored as the first crowding indicator. It can be found in Jenelius (2019) or in Li *et al.* (2013). The user's number does not represent correctly the crowding of a car. A large car can welcome more users onboard than a smaller one. Therefore Li *et al.* (2013) use $\frac{User'snumber}{Seat'snumber}$ or the user's number divided by the available area in m². Both indicators take into consideration the particularity of the car or the considered station.

Winter *et al.* (2018) consider the **occupancy rate** while building an Automated Demand Responsive Transport Service. This indicator is presented by Fielding *et al.* (1978) and can be computed from the passenger miles. Winter *et al.* (2018) maximize the occupancy rate of its vehicles in order to lower the needed fleet size. Li *et al.* (2013) reviewed different specification of crowding measures defined by transport authorities in different countries. They conclude that if the number of standing passengers per square meter can be an indicator for short journeys, the number of available seats should be the only capacity indicator for long journeys. They are also in favor of surveys to collect the perceived crowding by the users; these surveys enable the calibration of indicators according to a subjective point of view.

Jenelius (2018) takes into consideration that the more crowded a train is, the longer the travel seems for the user. Jenelius (2018) extends the Real Buffer Time Indicator by including the user's comfort. This comfort is linked to the number of passengers in the train and to the position (seating/standing) of each passenger. A queuing method is defined to deal with the passenger position and different multipliers are provided. They introduce the Experience Service Reliability Gap (ESRG), which is the difference between an upper percentile and the median perceived journey time for an OD. The perceived journey time is the actual travel time, waiting time and transfer time weighted by a time multiplier corresponding to the crowding conditions.

The **Demand Supply Gap** is introduced by Blandin *et al.* (2019) as the number of users who cannot board a train. They develop the FASTER algorithm to compute the DSG according to different sensors like CCTV cameras or turnstiles. The DSG enables the operator to see how crowded a train line is, and whether an increase of train frequency seems necessary. Bešinović *et al.* (2019) minimize the number of denied passenger in their disruption management strategy. Their strategy aims at minimizing the total passenger delays, the number of denied passenger and the adjustment to train service. They reschedule trains and control passenger flows when perturbations occur. They take station capacity into consideration and their strategy may recommend passenger to wait outside the station if it is overcrowded.

KPIs for freight transportation

Globally, the KPIs are quite the same than those explored in section 2.1. However, we can also focus on specific activities related to freight transportation. Thus, the selected KPIs are mainly based on the works of the following EU projects: FP7 C4R (DICEA, 2015) and on the OptiYard project (Shift2Rail, 2019), which selected a body of KPIs adapted to marshalling yards (MY) activities. These KPIs have been grouped into categories allowing the evaluation of the specific characteristics of MYs. With a qualitative assessment, the interactions between the MY and the surrounding network have been highlighted with the help of a heat map showing the impact of the surrounding network on the MY. The categories are reported in Table 1.

Table 1 KPI set for the assessment of the marshalling yard

Category	KPI	Unit	Description
Operational quality	Mean wagon transit time	[h]	Mean wagon transit time
	Mean wagon idle time	[h]	Time waiting for performing the next step/process
	Number of wagons sorted over the hump	[-]	Number of wagons sorted over the hump during a time interval
Yard capacity	Arrival yard utilization factor	[%]	Track length occupation/Sum of track length
	Classification yard utilization factor	[%]	Track length occupation/Sum of the track length
	Departure yard utilization factor	[%]	Track length occupation/Sum of the track length
	Number of wagons in the marshalling yard	[-]	Number of wagons in the marshalling yard at the same time
Operational competitiveness	Personnel needs	[h]	Sum of labour hours
	Rolling stock use rate	[km]	Sum of distance travelled by shunting locomotives
Operational reliability	Delays of outbound trains	[h]	Sum of Expected Time of Departure (ETD) delays
Operational resilience	Resilience	[h]	Sum of time operating under degraded mode

2.3 Material

The material should also be monitored by KPIs. The problem detection has a huge role to play in the forecasting of perturbation. Furthermore, the growing number of users makes safety more and more important to avoid dramatic accidents or huge delays.

2.3.1 Safety

This part is just about safety problems generated by the infrastructure or the rolling stock.

Famurewa *et al.* (2014) use the failure frequency to evaluate functional failure and infrastructure reliability. Failure frequency counts the number of failures which interrupt the traffic flow. Minor failures are not considered in the failure frequency "because of the extensive and complex nature of railway systems".

Jamshidi *et al.* (2017) study the detection of a rail defect called squat. They introduce a detection method to identify them. They also define KPIs to monitor the presence of squats according to their type. They also aggregate them in a fuzzy indicator to build a normal zone according to different scenarios. The squat growth is also studied to forecast the moment when the track could break. Their fuzzy method allows the observation of the worst-case scenario for the squat growth. Their prediction allows to plan maintenance works when they are necessary and before an impactful failure.

2.3.2 Perceived safety

Perceived safety is naturally linked to safety, but also to customers. Osswald *et al.* (2012) define perceived safety as "the degree to which an individual believes that using a system will affect his or her well-being". They qualify perceived safety as critical in the user's intention to use the system. The different categories used to evaluate how the user perceives safety go from "I believe that using the system is dangerous" to "I can use the system without looking at it". The authors explore also the field of anxiety along the perceived safety.

Nordhoff *et al.* (2020) work on the perceived safety and the interaction with automated shuttles. They distinguish 8 key-points for the perceived safety. Speed, dynamic object and event identification, longitudinal and lateral control, emergency button, trust in technology, automated shuttle sharing, controlled environment and the behavior of other road users affect the most the perceived safety of the users. Another part of the study focuses on interaction between the shuttle and the outside environment.

2.3.3 Autonomy

The vehicle's autonomy is an important concern for the decentralization and the swarm technology. Both rely on communication between vehicle and infrastructure and on their ability to make good decision according to the situation

The difficulty to qualify an autonomy's degree leads to monitor the number of autonomous vehicles at each moment. An autonomous vehicle is a vehicle which does not need to communicate with a control center to make decisions. This indicator could be linked to the distance travelled by autonomous vehicles described by Kloostra *et al.* (2019). Kloostra *et al.* (2019) study the impact of autonomous road vehicles according to their proportion in the travelling vehicles.

Nevertheless, the regulator must control the decision made by vehicles on field. The regulator's approval is important and should not be neglected. The KPI introduced by Park *et al.* (2012) count the number of interventions number done by the control center. If this KPI is used to check if an autonomous system needs help from the control center, it could easily be reversed for a control center to check if the on-field decision is good enough.

2.3.4 Public acceptance

Public acceptance is a mix between all the categories listed above. It can mainly be integrated thanks to the use of surveys or polls. Public acceptance plays an important role in transports' success. If the public does not want to use a public transport, it will decline.

Pakusch *et al.* (2017) investigate user's acceptance of fully automated public transport. They want to check on 3 hypotheses: Acceptance of fully automated public transport depends on age, Previous experience with autonomous vehicles increases acceptance of fully autonomous public transport and acceptance of fully automated rail-bound vehicles is greater than the acceptance of fully automated non-rail-bound vehicles. They conclude their study by saying that approximately 77% of the respondents can imagine using regularly autonomous public transport in the future.

Madigan *et al.* (2017) adapt the criteria described in the UTAUT from Vendakesh *et al.* (2012). They hold on to 5 criteria which are performance expectancy, effort expectancy, social influence, facilitating conditions and hedonic motivation. They apply their framework to an Automated Road Transport System's demonstration in Trikala (Greece). They show that the five criteria play a big part in the user's desire to use ARTS again.

2.4 Environment

The environment must be taken into consideration, nevertheless only the environment impacted by the transport system must be measured. So, three main environmental fields have emerged in the bibliography: noise pollution, pollutant emission and power consumption. The decision to focus on direct interaction between the transport system and the environment rely on the fact that the measurements must measure only decision related to the transport system. The selected KPIs must mainly depend on transportation policies and no on other urban policies.

2.4.1 Noise perturbation

Noise is a particular concern for transportation policies. Inhabitants living near railway lines may complain of high noise level. Jeon *et al.* (2005) give an overview of sixteen sustainable transportation initiatives, two of them mention noise. The TAC (Transportation Association of Canada) conducted a study in 1999 in Ottawa. This study encourages operator to "limit noise intrusion below levels accepted by communities". The other study is funded by the Centre for Sustainable Transportation (CST) and the government of Canada. The CST gives different points that a sustainable transportation system must respect. Among them can be found the minimization of "the use of land and the production of noise". Nevertheless, neither studies give a KPI to monitor noise. Hennino (2007) formulate indicators to get closer to the "ideal" indicators of sustainable development advocated by the Global Reporting Initiative (GRI). Therefore, she gives a noise KPI as the "**Percentage of population subjected to a predefined threshold**". She also monitors the number of rail kilometers treated with insulation system. The number of complaints related to noise has to be counted, as well as the average delays of response. These two last indicators are computed *a posteriori.* Haghshenas *et al.* (2012) collect different urban KPIs in various studies. The category noise pollution gathers four indicators with a frequency of use of 13. The authors then highlight the following indicator: "Population exposed to noise > 55dB". This indicator encourages the operator to monitor the noise level. Because of the impact of noise on the population, the measurement of the average noise level, the noise peaks and their frequency seem important.

2.4.2 Energy consumption

The energetic consumption remains an important point for the preservation of the environment and from an economical point of view. Nevertheless, its definition differs from an author to another. For example, Hennino (2007) evaluates not just the total consumption, but also the electricity consumption, its proportion in the total consumption, the electricity recovering during braking. A focus on fossil fuels is made, the energy brought by diesel, diester, steam or petrol is computed. The part of renewable energy is also mentioned. She also reports some KPIs on energy efficiency with the energy efficiency of traffic (passenger.km/kilo of oil equivalent) and the energy efficiency of service (available places * kilometer / kilo of oil equivalent). All energy KPIs are expressed, either in percentage or in ton of oil equivalent. Haghshenas et al. (2012)'s goal is to minimize consumption of non-renewable resources and to limit the consumption of renewable resources. They count three indicators for a frequency of use of 11 on energy consumption and four indicators for a frequency of use of eight for the category renewable energy type. They measure the transportation energy consumption in Megajoules. Cao et al. (2014) reschedule in real time the timetable according to a stochastic demand. They aim at optimizing the timetable's deviation, the passenger travel time and the active energy consumption defined by Wang et al. (2014). The active energy consumption depends on the vehicle's speed, its acceleration and the time it spends at cruise's speed. Mlinaríc et al. (2018) weight different KPIs for the evaluation of railway intelligent transport system with the Group Analytic Hierarchical Process.

Energy consumption is considered as "the most important in the assessment of railway ITS effects". Åhren (2005) uses the energy consumption per area to evaluate its consumption. Wang et al. (2010) prefer using only the "unsustainable energy per ton miles required" to compute the energy consumption. Their choice is motivated by the higher negative impact that have unsustainable energy on the environment.

2.4.3 Pollutant emission

The pollutant emission is also essential to study the environmental impact of a transport system. The literature gathers more than just one pollutant. Greenhouse gas (GHG), CO_2, NO_2 or fine particulate are all to be monitored according to different scientific articles. Wang et al. (2010) do not differentiate the emission of different pollutant. They aggregate them into a unique KPI: the pollutant released in tons of mobile pollutant per tons miles required. Mlinaríc et al. (2018) give to the amount of GHG emission the second highest weight to evaluate the impact of intelligent transport system. Nevertheless, they highlight a weakness of this indicator: the GHG emission can only be inferred from other indicators. Haghshenas et al. (2012) main goal is to « Limit emissions and waste within the planet's ability to absorb them ». Therefore, they develop indicators to monitor the pollutant emission. They monitor different local air pollutant like CO, VOC (Volatile Organic Compound), NO_x per capita and the emission from transports which is mainly CO_2-CH_4. They compare also different parts of the world according to transport and conclude that the developed part of Asia and Europe have the best composite sustainable transport index because an emphasis is made on non-motorized transport.

Sun et al. (2018) design an indicator system to evaluate the benefits brought by urban public transport infrastructure. They monitor the level of NO2, of SO2, of particulate matter and of air quality around the public transport infrastructure. They also say that the adoption of advanced technologies and the raising efficiency of energy use will reduce the emission of pollutant. Jeon et al. (2005)'s overview on sustainable transportation initiatives show that the reduction of pollutant, mainly GHG, VOC or CO, is taken into consideration. All initiatives but one monitor either the CO2 emissions or the GHG emissions. Some are even more precise and monitor the NOx, the VOC and the CO emissions.

Table 2 KPIs on railway management

Type / Category	Ratio	Speed	Time	Value	Euros	Frequency	Other
Generic	Capacity utilization*	Commercial speed*	-	#vehicles* Travelled kms per vehicle	Ticket price	-	Max customer flow per hour
Production	NRR*	-	Forecast of secondary delays* Forecast of train's late arrival*	#skipped stops* # changes* #impacted vehicles*	Cost adjustment*	-	-
Timetable	Regularity Index Systematic Timetable Index	-	Dwell time	-	-	Train frequency*	Headway regularity
Maintenance	-	-	Repair time or Interruption*	-	Cost of Maintenance Work*	-	-
Platooning	-	Convoy's speed*	-	# vehicles in a convoy* Size of each convoy*	-	-	Energy saved
Decentralization	-	-	Decision-making time*	PA/Auction value*	-	-	-
	-	-	Conflict resolution duration*	-	-	-	-

#: Number of	*: Realtime operator	IM/Regulator Company Both	Traveler	Urban community / Organizing authorities

Table 3 KPIs on Mobility

Type / Category	Ratio	Speed	Time	Value	Euros	Frequency	Other
Instantaneousness	Running ratio	$Euclidean\frac{distan}{TravelT}$	TT*	-	-	-	-
			In-vehicle TT*				
			AATT*				
			Dwell time*				
Immediacy	-	-	Waiting Time*	-	-	-	-
			Boarding/Alighting Time*				
			Transfer WT*				
			Idle Time*				
Reliability	Ratio of trips delayed more than 5 min	-	Cumulative delay*	#Vehicles above a delay's threshold *	-	-	Coefficient of variation
	Ratio of respected Headways		Average Punctuality				
			Range Variation				
			Excess WT*				
Comfort	$Users's\frac{numb}{Seat'sn}$	-	ESRG*	#users*	-	-	-
	Occupancy rate *	-	-	DSG*	-	-	-

#: Number of	*: Realtime operator	IM/Regulator Company Both	Traveler	Urban community / Organizing authorities

Table 4 KPIs on freight transportation

Type / Category	Ratio	Speed	Time	Value	Euros	Frequency	Other
Operational quality	-	-	Mean wagon transit time Mean wagon idle time	# wagons sorted over the hump	-	-	-
Yard capacity	-	-	-	# wagons in the marshal-ling yard Arrival yard utilization factor Classification yard utilization factor Departure yard utilization factor	-	-	-
Operational competitiveness	Rolling stock use rate	-	Personal needs	-	-	-	-
Operational reliability	-	-	Delays of outbound trains	-	-	-	-
Operational resilience	-	-	Resilience	-	-	-	-

#: Number of	*: Realtime operator	IM/Regulator Company Both	Traveler	Urban community / Organizing authorities

Table 5 KPIs on environment

Type / Category	Ratio	Speed	Time	Value	Euros	Frequency	Other
Noise	Ratio of population subjected to a predefined threshold	-	-	#complaints related to noise	-	Frequency of peak's noise	Average Noise Level* Peak noise
Energy	Part of renewable energy	-	-	Energy consumption Electricity consumption	-	-	Passenger.km/koe
Other	-	-	-	-	-	-	GHG/CO2/NO2 emission*

#: Number of	*: Realtime Operator	Regulator Company Both	Urban community / Organizing authorities

Table 6 KPIs on material

Type Category	Ratio	Speed	Time	Value	Euros	Frequency	Other
Safety	-	-	-	-	-	Failure frequency	Squat density Perceived safety
Autonomy	-	-	-	# autonomous vehicle *	-	-	-

#: Number of	*: Realtime operator	IM/Regulator Company Both	Traveler

3 OPERATIONAL PRINCIPLES

In order to decide what decision-making objectives or KPIs to be used in the self-organizing system, it is essential to establish first the operational principles to be followed. These principles have to be compliant with operational rules as much as possible, for at least two reasons: 1. To make its future deployment envisageable, and 2. To enable a fair comparison with a centralized TMS. To produce such compliant principles, our approach has consisted in interviewing several experts from Western-Europe infrastructure managers. We focus our attention on everyday perturbations and we exclude big disruptions. The answers are detailed in Appendix A and they are summarized in the next section.

3.1 Interview of the infrastructure management experts

To interview the experts (regulators/controllers/dispatchers), we have built a questionnaire around the following themes in particular: automation level in decision-making, circumstances for the decision-making, contingency plans, possible decisions in disturbed situation, available times for the decision-making, control areas and negotiation.

The interviewed experts work for the following infrastructure managers, exhaustively:

- Bahne Danmark (Urban network - Copenhagen)
- DB Netz (General network – Germany) – *Confidential answers*
- IP (General network – commando ferroviario centro, Portugal) – *Confidential answers*
- Network Rail (General network – GB)
- RFI (General network – Milano/Brescia/Verona)
- SNCF Réseau (General network – France).

3.1.1 Overall results

It is interesting to observe that most of the answers are quite homogeneous, which shows a great similarity in the decision-making process in the different countries:

- The control command is manual, and few semi-automatic (with manual acceptance) systems are deployed
- There is no definition of set of trains to be considered when solving a conflict
- Any area can make the object of traffic management decisions
- Decisions can be made and changed anytime as far as they can be implemented (for route setting or other aspects as passenger information)
- There is no contingency plans for perturbations
- The conflict definition is classic: it appears when theoretical free-network paths overlap.

3.1.2 Results about the used KPIs

Following the experts' advice and views, we can gather KPIs common to all railway networks (general) and distinguish some of them proper to urban railway networks, as listed hereafter:

- **General railway network:**
  - total delay / delay for trains above 3 min (mentioned as the prior KPI for the IMs)
  - passenger travel and waiting time
  - maximum delay, recovery time, number of late trains

  Other measurement performance scores:
  - Percentage or number of trains less than 5/6/10/15 min late at destination.
  - Total arrival delay at destination per type of service
  - Delay increase of more than 3 min
  - Percentage of passengers less than 15 min late (long distance)
- **Urban railway network**
  - passenger travel and waiting time
  - timetable recovery time
  - total delay, maximum delay, delay for trains above a threshold, number of affected trains,
  - number of actions taken,
  - number of broken transfer connections (if any)

3.1.3 Possible decisions to be taken

Also from the experts' answers, the decisions that can be made to deal with perturbations concern both the IM and the RUs. Precisely, the decisions made by the IM at two levels (local: dispatchers, network: controllers) are the following:

- Retiming
- Reordering
- Local rerouting
- use of track portions typically used for the opposite direction

Sometimes, some decisions are initiated by the RUs:

- skip or add stop
- cancelling or addition of a train
- preservation or breaking of transfer connections
- retiming, e.g., when a threshold is exceeded.

3.2 Principles of self-organization process

In the SORTEDMOBILITY project, we propose self-organizing principles to deal with everyday rail operations, i.e., **including traffic perturbations but excluding**

disruptions. According to these principles, traffic management works in a fully automatic way.

The performance of self-organization is compared with the one of an optimized centralized system, as well aiming at dealing with the everyday operations.

The definition of the overall traffic management process is out of the scope of SORTEDMOBOLITY. If suitable, it will be possible to design this process so that dispatchers always monitor the traffic evolution and are able to take over self-organizing trains whenever appropriate.

Specific actions such as manual overtaking will be triggered by particular traffic conditions. One may imagine dispatchers exploiting specific decision support tools, whose definition is out of the scope of this project.

Lack of consensus

In self-organised traffic, any plan relies on a consensus between the agents. If no consensus is reached, the applied real time traffic plan (RTTP) (Quaglietta *et al.*, 2016) will be the last accepted one. However, there are other approaches to explore and possibly compare, such as:

- RTTP imposing simple rules (e.g., priority to on-time trains),
- RTTP imposing 'first come first served' policy,
- RTTP as close as possible to the original timetable,
- best RTTP according to some criterion, found after computing different ones.

It is also possible to consider and design articulated consensus systems to link decisions made at different locations and times (e.g., token system). Approaches different from the retained one will possibly be considered as evolutions of the SORTEDMOBILITY system.

3.3 Self-organization vs centralization

3.3.1 Allowed train decisions in self-organized systems

Self-organizing trains can make all decisions, as they actually are RUs: retiming; reordering; local rerouting; use of track portions typically used for the opposite direction; skip or add stop; short-turning.

Centralized system can make decisions that do not need negotiation: retiming; reordering; local rerouting; use of track portions typically used for the opposite direction

3.3.2 Comparison of self-organization vs centralization

The comparison will be done considering two sets of decisions:

1. **A restricted set of decisions for self-organization**, equal to the one possible for centralized systems. This will allow the evaluation of the impact of self-organization and the verification of the following conjecture: if the control area is not too large for the centralized system, self-organization suffers from local vision and compromises needed for consensus.
2. **All decisions**. This will allow us to measure the impact change if further possibilities offered by self-organization are exploited, such as immediate stakeholder interactions. To do so, it is likely that new KPIs will be necessary.

4 DECISION-MAKING OBJECTIVES AND KPIS

In this chapter, we present the DMOs and KPIs chosen to drive and assess the self-organizing system for each of the use-cases proposed by the industrial partners of the SORTEDMOBILITY project. To choose the most appropriate ones, we discussed first with a panel of railway experts during a workshop (see appendix B). Then, the decision of DMOs and KPIS has been made within the consortium on the basis of all the elements at our disposal.

4.1 Decision-Making Objectives

As global aim, the system is intended to **minimize the overall delay, which can be seen as Decision-Making Objective**, but the delay cannot be considered equally whether it concerns trains, passengers, or cargo. There are several kinds of delay which cannot be added. Moreover, in self-organizing traffic, it is necessary that all the agent utilities share the same value unit because each agent computes its utility following its own function.

The chosen unit of utility value thus represents a cost associated to the delay, which may be expressed in Euro currency. It is then required to be able to convert any delay value into Euro on the basis of cost functions.

Concretely, we consider the following delay costs:

- Train delay: cost per lost second [€ / sec]
- Passenger delay: cost per lost second.passenger [€ / sec.passenger]. The more crowded the train, the greater the cost.
- Cargo delay: cost per lost second.t [€ / sec.t]. It has to be noted that a freight can transport several types of cargo which may be more or less precious or perishable. The cost may be thus a weighted average value or should be decomposed to be computed accurately.

From the above delay costs and the utility functions, it is possible to compute and sum all utility values since they share the same unit expressed in Euro.

In a competitive railway system, in which several RUs operate trains on the same network, Passenger and Cargo delay costs may be private information.

4.2 Choice of the Key-Performance Indicators

Differently to the Decision-Making Objectives whose number must not be excessive (1 or two objectives), it is possible to keep a large number of KPIs to monitor and assess the self-organizing traffic. Hence, the panel of fields which may be covered is possibly large as presented in Chapter 2.

Moreover, there is no time constraint as may be the real time for the decision-making in traffic management. So, it is possible to keep KPIs assessed in real time and others computed either after a period of time or in post-operations process.

For the needs of the SORTEDMOBILITY project, two main questions have to be answered for choosing the KPIs. The first concerns the capability of assessing or computing the KPIs under consideration with the EGTrain simulator, or at least the availability of the required data for doing that in parallel of the simulation.

The second question is about the relevance of the KPI with regards to the use-case and its ambitions. Indeed, the use-cases which are addressed in the project are very different whatever their size, their field, or the used rolling-stock.

Table 7 and Table 8 present a non-exhaustive set of the KPIs which may be relevant and/or computable for the use-cases. Hence, for each of them:

- We indicate whether the KPI is relevant or not (applicability or material: battery, diesel, electrified)
- A KPI may seem relevant even if it is not clear how to compute it, e.g., 'Neutrality regarding the users' would deserve to be precised so as to make it computable.

Finally, except the KPIs not explicitly marked as computable or relevant and those not applicable, all the others may be potentially monitored and assessed for the use-case under question.

Table 7 Key-Performance Indicators that can be used per use-case (1/2).

	Key-Performance Indicator	Computable	French use-case Small line	Italian use-case Freight line	Danish use-case Mass-transit
Travelers	Generalized travel time (travel, waiting and transfer times)	✓	✓	x (not applicable)	✓
	Felt travel time	✓	✓	x (not applicable)	✓
	Passenger punctuality (% passengers with <5min late at destination)	✓	✓	x (not applicable)	✓
	Passenger delay	✓	✓	x (not applicable)	✓
	Train frequency	✓	✓	x (not applicable)	
	Neutrality regarding the users whether their pick-up location or time	?	?	x (not applicable)	?
	Accessibility	?	✓	x (not applicable)	✓
	Number of broken connections/impacted passengers	✓	✓	x (not applicable)	✓
Train performance	Train travel times	✓	✓	✓	✓
	Train delays	✓	✓	✓	✓
	Train dwell time	✓	✓	✓	✓
Financial impact	Operational cost	✓	✓	✓	✓
	Revenue (RU – passenger fees, IM – track usage fees)	✓	✓	✓	✓
	Trading of unused freight paths between RUs	✓	x (not applicable)	✓	x (not applicable)

Table 8 Key-Performance Indicators that can be used per use-case (2/2).

	Key-Performance Indicator	Computable	French use-case Small line	Italian use-case Freight line	Danish use-case Mass-transit
Resource utilization	Energy consumption	✓	✓	✓	✓
	Travelled distance	✓	✓	✓	✓
	Line capacity utilization	✓	✓	✓	✓
	Number of rolling stock/materials/vehicles	✓	✓	✓	✓
Environment	Pollutant emission	✓	x (batteries)	diesel or electrified?	x (electrified)
	Noise emission	?	?	?	?
	Equity between RUs	?	x (not applicable)	✓	x (not applicable)
Organization	Time spent on communication/coordination	?	✓	✓	✓
	Time required for negotiations	?	✓	✓	✓
System	Reliability of the plan	?	✓	✓	✓
	Robustness of the solution	?	✓	✓	✓
	Resilience	?	✓	✓	✓
	Flexibility of the solution	?	✓	✓	✓
	Stability of the solution	?	✓	✓	✓
Safety	Passengers crossing tracks	?	x (not applicable)	x (not applicable)	✓
	Number of unscheduled stops at dangerous points (bridges), due to restrictive signals	✓	✓	✓	✓

5 FINAL SYNTHESIS

The work package 1 aims at defining the operational principles and the Decision-Making Objectives which will drive the self-organizing traffic, and at proposing Key-Performance Indicators which will be monitored and assessed.

To do so, in Chapter 2, we have reviewed many indicators in the literature related to the ground transportation fields. These indicators must enable the evaluation of the three main points needed to build a new transport system fairer, easier to use and which excludes the smallest possible population share. Therefore, these indicators enable the evaluation of levels of service, of accessibility and efficiency of transport service. The literature from road transportation provides several indicators interesting for the evaluation of platooning and for the autonomous transport systems. All the collected indicators have been sorted into different categories to specify the operational fields they provide information to. Depending on their interests, different stakeholders may use indicators related to specific operational fields and different time-horizons (from real-time to post-operations time).

In parallel to the review on indicators, we have defined the operational principles which will be at the basis of the self-organizing traffic. It has been done from interviews with railway infrastructure managers who are expert in traffic management and dispatching, as related in Chapter 3. It results from these interviews a large consensus among the European IMs on traffic management principles. Delay appears the main issue, whatever the way of assessing and minimizing it.

Indicators were gathered considering two goals: 1. drive and 2. monitor the transport system. We have detailed in Chapter 4 the approach and the workshop organized to discuss what should be the DMOs to drive the self-organizing traffic and the KPIs to monitor and assess it. As for DMOs, we decided to convert into Euros any kind of delays to be considered: the delays hence monetized correspond to a cost which has to be minimized. The advantage is to be able to manipulate together delays which do not characterize the same things (passengers, cargo) or may not concern the same stakeholders (travelers, RUs, IMs, authorities). The choice of the KPIs is much more open and depends essentially on the availability of the required data and on the relevance of the KPI with regards to the use-case under consideration.

6 REFERENCES

[1] T. Albrecht, A. Binder and C. Gassel, "Applications of real-time speed control in rail-bound public transtransport systems," *IET Intelligent Transportation Systems,* 2012.

[2] S. Sobieraj, J. Rodriguez and G. Marliere, "Simulation of solution of a fixed-speed model for the real-time railway traffic optimisation problem," in *4th International Seminar on Railway Operations Modelling and Analysis RailRome*, 2011.

[3] B. Jaekel, T. Albrecht and F. Thonig, "Comparative Analysis of Algorithms for Train Running Simulation," in *5th International Seminar on Railway Operations Modelling and Analysis RailCopenhagen2013*, 2013.

[5] S. Osswald, D. Wurhofer and S. Trösterer, "Predicting Information Technology Usage in the Car: Towards a Car Technology Acceptance Model," in *International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 2012.

[6] H. Haghshenas and M. Vaziri, "Urban sustainable transportation indicators for global comparison," *Ecological Indicators,* pp. 115-121, 2012.

[7] R. Cascajo and A. Monzon., "Assessment of innovative measures implemented in european bus systems using key performance indicators," *Public Transport,* pp. 257-282, 2014.

[8] F. Antonialli and e. al., "International benchmark on experimentations with autonomous shuttles for collective transport," Centrale-Supélec, 2019.

[9] T. Åhren, "A study of maintenance performance indicators for the Swedish railroad system," Lulea tekniska universitet, 2005.

[10] L. Eboli and G. Mazzulla., "Performance indicators for an objective measure of public transport service quality," *European Transport,* 2012.

[11] R. M. Goverde, "Railway timetable stability analysis using max-plus system theory," *Transportation Research Part B: Methodological,* pp. 179-201, 2007.

[12] X. Delorme, X. Gandibleux and J. Rodriguez., "Stability evaluation of a railway timetable at station level," *European Journal of Operational Research,* pp. 780-790, 2009.

[13] I. A. Hansen, "Railway network timetabling and dynamic traffic management," *International Journal of Civil Engineering,* pp. 19-32, 2010.

[14] G. Georgiadis, Y. Xenidis, I. Toskas and P. Papaioannou., "A performance measurement system for public transport services in thessaloniki, greece," in *Transport Research Arena (TRA) 5th Conference: Transport Solutions from Research to Deployment*, Paris, 2014.

[15] H. Orth, N. Carrasco, M. Schwertner and U. Weidmann., "Calibration of a public transport performance measurement system for switzerland," *Transportation Research Record,* pp. 104-114, 2013.

[16] L. Sen, M. Highsmith, L. Cherrington, C.Weatherby, S. R. Majumdar and e. al., "Performance measures for public transit mobility management," Dept. of Transportation. Research and Technology Implementation Office, Texas., 2011.

[17] B. Djordjevic and E. Krmac, "Key performance indicators for measuring the impacts of its on transport," in *ISEP*, 2016.

[18] T. J. Mlinaríc, B. Djordjevíc and E. Krmac., "Evaluation framework for key performance indicators of railway its," *Promet Traffic & Transportation,* pp. 491-500, 2018.

[19] B. Schittenhelm and A. Landex., "Danish key performance indicators for railway timetables," in *Annual Transport Conference*, Aalborg University, 2012.

[20] F. Proboste, J. C. Muñoz and A. Gschwender., "Comparing social costs of public transport networks structured around an open and closed brt corridor in medium sized cities," *Transportation Research Part A: Policy and Practice,* pp. 187-212, 2020.

[21] A. D'Ariano, *Improving real-time train dispatching: models, algorithms and applications,* Delft University of Technology, 2008.

[22] E. Quaglietta, P. Pellegrini, R. M. Goverde, T. Albrecht, B. Jaekel, G. Marlière, J. Rodriguez, T. Dollevoet, B. Ambrogio, D. Carcasole and e. al., "The on-time real-time railway traffic management framework: A proof-of-concept using a scalable standardised data communication architecture," *Transportation Research Part C: Emerging Technologies,,* pp. 23-50, 2016.

[23] X. Wu, H. Dong, C. K. Tse, I. W. Ho and F. C. Lau., "Analysis of metro network performance from a complex network perspective," *Physica A: Statistical Mechanics and its Applications,* pp. 553-563, 2018.

[24] A. Reddy, A. Cramer, J. Cucarese, A. Lu and M. Tran., "Performance measurements on mass transit-new york city transit authority case study," *Transportation Research Record,,* pp. 125-138, 2009.

[25] M. Marcelli and P. Pellegrini, "Literature review toward decentralized railway traffic management," *IEEE Intelligent Transportation Systems Magazine,,* p. 18p, 2020.

[26] M. Lewandowski, B. P. laczek and M. Bernas, "Decentralized control of traffic signals with priority for ambulances," *Journal of Medical Informatics & Technologies,* p. 26, 2017.

[27] N. Stojadinovìc, B. Boškovìc and D. Trifunovìc, "Train path congestion management: Using hybrid auctions for decentralized railway capacity allocation," *Transportation Research Part A: Policy and Practice,* pp. 123-139, 2019.

[28] G. Vernazza and R. Zunino., "A distributed intelligence methodology for railway traffic control.," *IEEE Transactions on Vehicular Technology,* pp. 263-270, 1990.

[29] P. Kavathekar and Y. Chen, "Vehicle platooning: A brief survey and categorization.," *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference,,* pp. 829-845, 2011.

[30] J. N. Yasin, S. A. Mohamed, M.-H. Haghbayan, J. Heikkonen, H. Tenhunen and J. Plosila., "Navigation of autonomous swarm of drones using translational coordinates," *International Conference on Practical Applications of Agents and Multi-Agent Systems,* pp. 353-362, 2020.

[31] C. Stenström, A. Parida and D. Galar., "Performance indicators of railway infrastructure," *The international Journal of railway technology,* pp. 1-18, 2012.

[32] G. Dupuy, L'urbanisme des réseaux, théories et méthodes., Armand Colin, 2011.

[33] M. M. Nesheli, A. A. Ceder and R. Brissaud, "Public transport service-quality elements based on real-time operational tactics," *Transportation,* pp. 957-975, 2017.

[34] Z. Cao, A. A. Ceder and S. Zhang., "Real-time schedule adjustments for autonomous public transport vehicles.," *Transportation Research Part C: Emerging Technologies,* pp. 257-282, 2014.

[35] E. Krmac and B. Djordjevíc., "An evaluation of indicators of railway intelligent transportation systems using the group analytic hierarchy process," *Electronics Science Technology and Application,* 2017.

[36] M. M. Nesheli, A. A. Ceder and T. Liu, "A robust tactic-based real-time framework for public-transport transfer synchronization," *Transportation Research Procedia,* pp. 246-268, 2015.

[37] R. Basu, A. Araldo, A. P. Akkinepally, B. H. N. Biran, K. Basak, R. Seshadri, N. Deshmukh, N. Kumar, C. L. Azevedo and M. Ben-Akiva, "Automated mobility-on-demand vs. mass transit: a multi-modal activity-driven agent-based simulation approach.," *Transportation Research Record,* pp. 608-618, 2018.

[38] M. Trompet, X. Liu and D. J. Graham., "Development of key performance indicator to compare regularity of service between urban bus operators," *Transportation research record,,* pp. 33-41, 2011.

[39] W. Ren, R. W. Beard and E. M. Atkins, "Information consensus in multivehicle cooperative control," *IEEE Control systems magazine,* pp. 71-82, 2007.

[40] B. Hennino, "Study of the problematic of the indicators of sustainable development for the French company RATP," 2007.

[41] Y. Wang, B. De Schutter, V. d. B. T.J., B. Ning and T. Tang, "Efficient bilevel approach for urban rail transit operation with stop-skipping," *IEEE Transactions on Intelligent Transportation Systems,* 2014.

[42] H. Wang and M. Jim, "A transportation performance measurement system with a mississippi case study," *Journal of the Transportation Research Forum,* 2010.

[43] C. M. Jeon and A. Amekudzi-Kennedy, "Addressing sustainability in transportation systems: de," *Journal of infrastructure systems,* pp. 31-50, 2005.

[44] W. Van Heeswijk and H. La Poutré, "Scalability and performance of decentralized planning

in flexible transport network," *IEEE,* pp. 292-297, 2018.

[45] N. Van Oort, "Incorporating enhanced service reliability of public transport in cost-benefits," *Public Transport,* pp. 143-160, 2016.

[46] N. Van Oort and R. Van Nes, "Regularity analysis for optimizing urban transit network design," *Public transport,* pp. 155-168, 2009.

[47] S. M. Famurewa, C. Stenström, M. Asplund, D. Galar and U. Kumar, "Composite indicator for railway infrastructure management," *Journal of Modern Transportation,* pp. 214-224, 2014.

[48] M. Lüthi, *Improving the efficiency of heavily used railway networks through integrated real-time rescheduling,* ETH Zürich, 2009.

[49] N. Van Oort, *Service reliability and urban public transport design,* 2011.

[50] E. Jenelius, "Data-driven metro train crowding prediction based on real-time load," *IEEE Transactions on Intelligent Transportation Systems,,* pp. 2254-2265, 2019.

[51] Z. Li and D. A. Hensher, "Crowding in public transport: a review of objective and subjective measures," *Journal of Public Transportation,* p. 6, 2013.

[52] E. Jenelius, "Public transport experienced service reliability: Integrating travel time and travel condition," *Transportation Research Part A: Policy and Practice,* pp. 275-291, 2018.

[53] S. Blandin, L. Wynter, H. Poonawala, S. Laguna and B. Dura, "Faster: Fusion analytics for public transport event," 2019.

[54] A. Jamshidi, A. Núñez, R. Dollevoet and Z. Li, "Robust and predictive fuzzy key performance indicators for condition-based treatment of squats in railway infrastructures," *Journal of Infrastructure Systems,* 2017.

[55] B. Kloostra and M. J. Roorda., "Fully autonomous vehicles: analyzing transportation network performance and operating scenarios in the greater Toronto area, canada," *Transportation planning and technology,,* pp. 99-112, 2019.

[56] R. Madigan, T. Louw, M. Wilbrink, A. Schieben and N. Merat, "What influence the decision to use automated public transport? using utaut to understand public acceptance of automated road transport systems.," *Transportation research part F: traffic psychology and behaviour,* pp. 55-64, 2017.

[57] H.-S. Park and N.-H. Tran., "An autonomous manufacturing system based on swarm of cognitive agents," *Journal of Manufacturing Systems,* pp. 337-348, 2012.

[58] C. Pakusch and P. Bossauer, "User acceptance of fully autonomous public transport.," *ICE-B,* pp. 52-60, 2017.

[59] I. Kaparias, M. Bell and M. Tomassini., "Using key performance indicators for traffic management and intelligent transport systems as a prediction tool," *ISIS,* 2011.

[60] H. Tu, H. Van Lint and V. Zuylen, "The influence of road geometry on travel time variabil-

ity," in *Proceedings of the 3rd international symposium on transport network reliability*, 2007.

[61] . Bešinović, Y. Wang, S. Zhu, E. Quaglietta, T. Tang and R. M. Goverde, "Integrated train and passenger disruption management for urban railway lines," in *IEEE Intelligent Transportation Systems Conferenc*, 2019.

[62] C. Bouvet-Agnelli, "Optimisation de la grille horaire Transilien du point de vue des voyageurs," 2016.

[63] M. M. O. Sidi, S. Hammadi, S. Hayat and P. Borne., "Urban transport network regulation and evaluation: A fuzzy evolutionary approach," *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans,* pp. 309-318, 2018.

[64] G. J. Fielding, R. E. Glauthier and C. A. Lave, "Performance Indicators for Transit Management," *Transportation 7,* pp. 365-379, 1978.

[65] M. G. Karlaftis and D. Tsamboulas., "Efficiency measurement in public transport: Are finding specification sensitive ?," *Transportation Research Part A: Policy and Practice,* pp. 392-402, 2012.

[66] P. Pellegrini, "Self-Organized Rail Traffic for the Evolution of Decentralized MOBILITY (SortedMobility)," 2021.

[67] O. Karoui, M. Khalgui, A. Koubâa, E. Guerfala and Z. L. E. Tovar., "Dual mode for vehicular platoon safety: Simulation and formal verification," *Information Sciences,* pp. 216-232, 2017.

[68] M. A. Khan and L. Boloni, "Convoy driving through ad-hoc coalition formation," *IEEE,* pp. 98-105, 2005.

[69] L.-M. Kieu, A. Bhaskar and E. Chung, "Public transport travel-time variability definitions and monitoring," *Journal of Transportation Engineering,* 2015.

[70] S. Nordhoff, J. Stapel, B. v. Arem and R. Happee, "Passenger opinions of the perceived safety and interaction with automated shuttles: A test ride study with 'hidden' safety steward.," *Transportation research part A: policy and practice,* pp. 508-524, 2020.

[71] S. Payne and H. Daniel, "Key performance indicators for intelligent transport systems," *in support of the implementation of the EU legislativ framework on its (Directive 2010/40/EU),* 2015.

[72] Q. Perrachon, R. Chevrier and P. Pellegrini., "Experimental study on the viability of decentralized railway traffic management.," *WIT Transactions on The Built Environment,* pp. 337-344, 2020.

[73] B. Su, P. Andelfinger, D. Eckhoff, H. Cornet, G. Marinkovic, W. Cai and A. Knoll., "An agent-based model for evaluating the boarding and alighting efficiency of autonomous public transport vehicles," *International Conference on Computational Science,,* pp. 534-547, 2019.

[74] Y. Sun and Y. Cui, "Evaluating the coordinated development of economic, social and environmental benefits of urban public transportation infrastructure: Case study of four chinese autonomous municipalities," *Transport Policy,* pp. 116-126, 2018.

[75] K. Winter, O. Cats, G. Correia and B. v. Arem., "Performance analysis and fleet requirements of automated demand-responsive transport system as an urban public transport service," *International journal of transportation science and technology,* pp. 151-167, 2018.

[76] F. Corman, A. D'Ariano, A. D. Marra, D. Pacciarelli and a. M. Samà., "Integrating train scheduling and delay management in real-time railway traffic control," *Transportation Research Part E: Logistics and Transportation Review,* pp. 213-239, 2017.

[77] B. Adeline, P. Dersin, E. Fabre, L. Hélouët and K. Kecir, "An efficient evaluation scheme for KPIs in regulated urban train systems.," in *RSSRail*, Pistoia, 2017.

[78] N. Van Oort, "Incorporating service reliability in public transport design and performance requirements: International survey results and recommendations," *Research in Transportation Economics,* pp. 92-100, 2014.

A. Infrastructure managers’ answers to the interviews[1]

	BDK	NR	SNCF	RFI
Type of network	urban (Copenhagen S-Bane)	a general railway network (mixed traffic with, e.g., long distance passenger trains, freight trains) (GB national rail network)	general railway network	general railway network (Milano-Brescia-Verona)
Who or what is in charge of traffic management decisions?	dispatcher for perturations and service manager for disruptions	dispatcher Some limited functions are automated (e.g. options for reordering trains and replatforming. These are identified automatically but the dispatcher still takes responsibility for implementing these decisions.)	dispatcher	dispatcher and automatic system: Dispatcher supervises the operation of the automatic system and can independently decide if and when operating manually

1 Deutsche Bahn and Infraestruturas de Portugal has also contributed, but the interview responses are confidential and are hence excluded from this report.

In what circumstances a traffic management decision is considered necessary? **a) when at least one train has a delay** **b) when a conflict is detected by the system** **c) periodically at constant intervals** **d) when unplanned high passenger flows emerge** **e) other**	all	b) with conflict defined as the competition for the resources. This includes: Headway, Junction margin, Junction crossing, Platform occupation etc. Consideration is a term used to describe an event that has the potential to affect train running but is not considered to be a conflict. This includes Turnaround Time, Traction type, Stock and Crew availability and Infrastructure restrictions, etc	a with delay larger than 5 min, b	a with a delay of 5/15/30 mintues depending on train catefory; B with a conflict is created between only two trains (without considering further propagation to others), i.e. when the deviation of the scheduled time of one train overlaps another train. Dispatcher must resolve the conflict (system only suggest but cannot resolve it automatically). Dispatcher can decide if works in manual or not

Are different decision making approaches used for different circumstances?	always human	- Some TMSs offer more localized approach which is providing conflict resolution in a small scale without considering the impact on a larger scale. Only one local solution is proposed. - Some other TMSs offer an approach that consider trains over a larger proportion of their journeys: depending on the system, some are able to identify all conflicts and simulate possible outcomes of interaction with the forecast. if a conflict is detected it will go through whether there are any resolutions rules available. It will propose the outcome of each to the operators. The operator can also simulate the outcome of each and select the most appropriate one for that conflict.	always human	Dispatcher, generally, works in manual in case of high traffic or in case of big perturbation.

How much time is available to make traffic management decisions?	any	dependsProcessing time goals would be dependent on the size and impact of the conflict and proposed solution; a decision that has only a local impact to one or two train services should be available quickly; if there is a significant event that requires many re-routing decisions, if the window of opportunity to interact is sufficient, it could be better to wait for a short amount of time for a better solution. For example, some routes have a restriction of a minimum two-hour window to be applied for conflict detection	Traffic management decisions should be made until the itinerary is set	If dispatcher works in manual can take the time he need for traffic management decisions.
To which extent the available decision making time is affected by the type of perturbation to be solved?	not really affected	-	No formalized decision making time. It depends on the context, one shall pay attention to the type of service, to elements as rolling stock and drivers, platforms, passengers	If the decision isn't urgent, a greater impact will be on the propagation of the disturbance
How does the available decision making time differ for	only one process	-	only one process	If automatic: dispatcher can only make decision before the event

different approaches?				
How are the trains and the area to be considered in a traffic management decision identified?	All trains impacted directly or indirectly by a decision are taken into consideration, in general	Depending on the type of TMS: only one or two trains or all trains in the area of control	Manual conflict prediction is based on done considering trains in neighboring control areas (zone d'éclairage) or up to train final destination for small lines, and big stations and junctions for long distance trains In the future, the whole French network will be considered, and even the European one	All trains impacted directly or indirectly by resolution conflict are taken into consideration, in general. But the system considers conflict for pairs of trains. Only after resolution of a conflict, another is generated.

Are particular areas of the network pre-identified for possible traffic management decisions in case of perturbation or is the whole network potentially involved?	the priority is to regulate the traffic outside of the central common section	There are local models (this will depend on the local operating model; most small conflicts that cause minimal disruption to the train service will not require the input of the Railway Undertakings (RU). Anything that significantly impacts on the RU or requires them to make a decision for their own operations, e.g. fleet or crew, will require RU involvement)	There are Controlled areas and Not Controlled areas, where there is little impact of conflict resolution decisions. The relevance of Controlled areas decreases when the distance from the most congested area (Paris) increases	Each dispatcher has its own jurisdiction over which it can make choices for the regulation / management of conflicts. There is another figure who can advise / propose choices on the whole network.
What decisions can be made? **a) retiming** **b) reordering** **c) local rerouting** **d) use of track portions typically used for the opposite direction** **e) skip stop** **f) additional stop** **g) additional train** **h) train canceling** **i) preservation or breaking of transfer connections** **j) other, please specify**	all, h seldom (j instead), g seldom	a b c E to I are taken with negotiation with Rus	A b c d e f (the rest are decisions of the RU) Rerouting options are dependent on the driver Local rerouting is in stations, in principle the RU negotiates but they always agre in normal situations.	All of the above. Letter e-f-g-h-i by service manager, a-b-c-d by dispatcher.

For traffic management decisions made for the types of perturbation we are dealing with, is there the need for a negotiation with train operating companies?	no negotiation for b c d e f Discussion for h and I (if any) Negotiation of a g	no negotiation a b c Negotiation e f g h I	no negotiation for a b d e (non-commercial stops) f (non-commercial stops) Discussion for h Negotiation for c e (commercial stops) f (commercial stops)	Letter "e-f-g-h-i" by service manager in agreement with RU. Letter "a-b-c-d" by dispatcher independently For c, discussion with RU in particular situations, as , e.g., if there is a passenger with reduced mobility on board For I, RU proposes but service manager can accept or refuse
What objectives are aimed at? **a) minimization of total delay** **b) minimization of maximum delay** **c) minimization of passenger travel and waiting time** **d) minimization of timetable recovery time** **e) minimization of delay only for trains more delayed more than … minutes** **d) minimization of the number of affected trains** **f) minimization of the number of actions taken** **g) minimization of the number of broken transfer connections** **h) other, please specify**	C has top priority; D comes second; All the others objectives are aimed at thereafter, in no particular order, according to the actual perturbation	a	a Frequence maximization on very busy lines for local passenger trains A simple and robust solution (from passenger perspective) is preferable	for e), 5 min for regional trains, 30 min for freight trains, 15 min for long trip trains

What measurements performance scores are based on the traffic management?	-	- Delay minutes (particularly for passenger trains) - The public performance measure (PPM) shows the percentage of trains which ran their entire planned journey calling at all scheduled stations and arriving at their terminating station within 5 minutes (for London & South East and regional services) or 10 minutes (for long distance services). It combines figures for punctuality and reliability into a single performance measure. It is the current industry standard measurement of performance	delay, local measures (wasted time due to an incident)	train arrival delay at the final station
How are they ranked?		-		

How many times decisions can be changed? And how much in advance final decisions must be made?	decisions can be made up and changed until last second	Route-specific	a b d up to the last feasible time C depends on the driver knowledge F g up to 10 minutes in advance (braking anticipation + stations noticed + station ready, e.g. platforms) A can be changed 3 times max For local rerouting in a station with a train stop, the decision must be made in advance to inform passengers	no limit
Are they contingency plans for perturbations?	mostly regarding what the service level (trains per hour) will be at each station: how many trains will be used for line, and where should they be turned	no	only for disruptions, including train canceling, passenger rerouting, alternative mode choice	There are contingency plans in the case of events / interruptions impacting traffic that presuppose planned traffic restrictions. In addition, there are contingency plans to be applied in cases of sudden great perturbation.

B. Workshop with academic and industrial railway experts

In addition to the KPIs review and the interviews of the railway infrastructure managers, the advices of railway experts have been collected during an online workshop involving more than 40 experts from RUs, IMs and railway engineering companies coming from many Western-Europe countries.

The experts were invited to answer two questions, both corresponding to a sequence of the workshop:

1. What may we get from self-organizing traffic?
2. What indicators shall be considered?

For each question, the answers were given according to the concerned stakeholder: the user, the RU or IM, the system.

Table 9 is the verbatim report of the answers to question 1, which have been only gathered into categories. Table 10 is the same for question 2, but with a distinction between KPIs and DMOs.

Table 9 Main expectations and risks expressed by the experts during the workshop.

USERS

ADVANTAGES

Optimised Travel time
- Travel time optimized and reduced in the transport system
- Better reaction time reducing travel time

Quality of service
- Higher frequency
- Accessibility
- Customised train trips: OD and time windows
- Accessibility for all groups of travelers (not clear)

RISKS

Planning and timetabling
- Bad estimation of arrival times
- Timetables difficult to be followed due to changing during the journeys
- Planning difficult to be done because of the absence of timetables.
- Delay and disruption possible because of several decision levels

Quality of service
- Quality of service degraded due to the flexibility leading to a loss of utility for certain passengers
- Traffic information confused

RAILWAY UNDERTAKINGS AND INFRASTRUCTURE MANAGERS

ADVANTAGES

Capacity
- Usage of capacity maximized
- Reduce high impact movements in the infrastructure

Human factors
- Reduced effect of human behaviours in dispatching
- Dispatchers and controllers could focus more on the parts of the network with high traffic.

System reactivity and availability
- Faster decisions and process. System reactivity increased.
- Direct communications of the trains among each other, bypassing the centralized system.
- Unmanned operations eliminate constraints related to the personnel availability

Costs
- Operational cost reduced for the IM BUT likely increased for RU
- Potential reduction in maintenance of infrastructure
- Economical operations needing less human intervention

Interactions, competition and cooperation
- No discrimination between RU
- Reorganisation of the interaction between RUs (more cooperation than competition)
- Reduction in necessary communications between RU and IM

Opportunity for tests
- A fully autonomous system could be used as a digital twin with real-time dispatching

RISKS AND QUESTIONS

Decision-making
- Bad decisions
- High complexity when (re-)planning rostering plans for rolling stock -> empty rolling stocks ?
- Actions and responsibilities from RU and IM with regards to the regulatory conditions
- Harmonization and Relationship between SO decisions and wider decisions from the centre.
- (Indirect) Role of the infrastructure in the decision-making process

Costs
- Higher Acquisition and maintenance cost of trains

SYSTEM

ADVANTAGES

Robustness and resilience
- Potential reduction of instances of conflicts
- Potentially more resilience
- Better approximation of optimality in practical cases

Connections and transfers
- Better integration with other transportation modes (including autonomous road vehicles)

RISKS

Rules and priorities
- National legislation and rules vs the required level of flexibility
- Emergence of RUs Inequalities?
- Risk for manipulation by clusters of users

Objective definition
- Potential conflict between specific objectives and overall objective of the system
- Scaling a decentralized system may be different from a centralized one.
- Not effective delay mitigation if coordination between agents is not well performed

Timetabling
- Uncertainty about timetables (definition and respect)

Communication failure

Table 10 KPIs and decision-making objectives for self-organizing traffic

USERS

Assessment KPIs

- Mobility
  - Travel, waiting and transfer time
  - Accessibility:
    - Flexibility both in time and space
    - Coverage: possible transfers
    - Frequency/number of departures per time unit
- Reliability of the travel plan -> punctuality, variability
- Neutrality regarding the users whether their pick-up location or time -> passengers weighting

Decision-making objectives

- Mobility :
  - Travel, waiting and transfer time
  - Passenger punctuality
- Number of planned and unplanned transfers
- Number of served destinations for freight trains
- Safety: avoid passengers crossing tracks

RAILWAY UNDERTAKINGS AND INFRASTRUCTURE MANAGERS

Assessment KPIs

- Resource utilization
  - Line capacity
  - Rolling stock
  - Energy consumption
  - Travelled km per vehicle
- Traffic management
  - Bias avoided in TM decisions
  - Number of trains crossing a bottleneck
- Operational costs for RU/IM (fuel, crew, maintenance)
- Railway system performance
  - Train travel time
  - Train dwell time
  - Train punctuality
  - Train regularity
- Trains delays

Decision-making objectives

- Resource utilization
  - Line capacity
  - Rolling stock
  - Energy consumption
  - Travelled km per vehicle
- Financial impact
  - Operational costs for RU/IM (fuel, crew, maintenance)
  - Fees : RU – passenger fees, IM – track usage fees
  - Efficient trading of unused freight paths between RUs
- Resource planning
  - Consistency with crew planning
- Trains delays
  - Overall delay for each solution
  - Number of trains impacted (slowed) by a decision
- Safety
  - Min Number of unscheduled stops at dangerous points (bridges), due to restrictive signals
- Equity
  - guarantee fairness in RU competition by fair priority management of the trains

SYSTEM

Assessment KPIs

- Organization
  - Time spent on communication/coordination
  - Time required for negotiations
  - Quality of the solutions compared to those obtained by a centralized TMS
  - Sustainability / attracting freight users (flexible and short-term reactions))
  - Adherence to priority rules given by the authorities
- Passenger punctuality
- Robustness / stability of the solutions

Decision-making objectives

- Environment
  - Minimise pollutant and noise emission
- Minimise energy consumption
- Maximise matching between supply and demand in a given time window -> several possible KPIs
- Accessibility given to low-demand areas
- Deadlock avoidance