Economic Steering Group

Task force on registers and databases

Final Report

EUROPEAN UNION
AGENCY
FOR RAILWAYS

120 Rue Marc Lefrancq | BP 20392 | FR-59307 Valenciennes Cedex
Tel. +33 (0)327 09 65 00 | era.europa.eu | @ERA_railways

Executive Summary

The Economic Steering Group (ESG) Network Framework envisages the organisation of topic-based Task Forces in order to collect evidence on the outcomes of the various objectives from the Agency's work programme. The ESG Task Force (TF) on Registers and Databases was set up to help aligning the Agency's work in the area of registers and databases with the business needs of the railway operators and other concerned stakeholders. The TF consisting of 11 experts nominated by organizations from across the EU facilitated by an external rapporteur (Vaibhav PURI, Rail Safety and Standards Board) and an internal co-rapporteur (Vojtech EKSLER, ERA) have over five months systematically analysed business use cases of RUs related to data from other parties, while using standard business analysis and data management techniques, in conjunction with economic evaluation approaches.

The TF took a holistic approach and developed a comprehensive methodology for the identification of data needs and for prioritization of improvements in their central availability and applied it to the use cases under five phases of train service provision. These were further refined and validated by a poll of seven RUs, external to the TF members, who have provided comments and inputs to the TF analytical work.

A series of high-level data models were developed to determine an ideal data architecture aligned with the business needs of the RUs and a fitness check was carried out to compare how it fits with the current architecture of registers maintained by the Agency.

On this basis, 21 improvement cases for some form of collection, management and public provision of centralized data were identified. They were then subject to qualitative (all) and quantitative (four selected cases) economic assessment (which were also used to validate the qualitative assessment). In general, use cases relying on better/more infrastructure-related data were considered as the most beneficial to the RUs, with urgent need to assure that current and any proposed approaches are fit to the purpose to meet the stated objectives (e.g. vehicle-route compatibility check). To this end the TF also recommended the need to create a strategic vision for the use of such data to enable and facilitate the implementation of future rail services and associated technological and operational solutions which are considered beneficial for the rail sector. This will allow the registers and their evolution to match the vision of the railway to ensure they are and remain fit for purpose.

The various use cases (either defined in legal framework or stemming from business needs) also call for a renewed and better focus on comprehensive data management, with the focus mainly on regulatory/administrative data management, which are within the scope of Agency activities. Key roles for the Agency were identified as a reference (railway) data authority and a data quality promoter. In these roles, any agency activities should promote and facilitate consistency, coherence and timeliness of data provision that meet both business and customer needs.

Based on the findings, the TF has identified a series of specific recommendations under vision/target system, Improvements in existing databases, new databases developments, EU railway data management and EU legal framework. If implemented and considered, they have a potential to steer Agency work programme and even Commission activities in the area of shared railway data, with the ultimate goal to facilitate the daily business operations of European Railway Undertakings.

Contents

1. General references, definitions and abbreviations

1.1. Reference Documents

	Title	*Reference*	*Version*
1.	Regulation (EU) 2016/796 of the European Parliament and of the Council of 11 May 2016 on the European Union Agency for Railways and repealing Regulation (EC) No 881/2004	Agency Regulation	11 May 2016
2.	Agency Economic Steering Group (ESG) Network Framework	ESG Network Framework	2 June 2017
3.	RINF Decision (Decision 2014/880/EU)	RINF decision	26 November 2014
4.	Commission Implementing Decision 2011/665/EU on the European register of authorised types of railway vehicles (ERATV)	ERATV decision	04 October 2011
5.	Agency Project Plan “Economic Steering Group - Registers and Databases Task Force”	TF plan	14 November 2018
6.	Directive 2012/34/EU establishing a single European railway area (recast)	SERA Directive	21 November 2012
7.	Directive (EU) 2016/797 on the interoperability of the rail system within the European Union	Interoperability Directive	26 May 2016
8.	Directive (EU) 2016/798 on railway safety	Railway Safety Directive	26 May 2016

1.2. Definitions and Abbreviations

1.2.1. General definitions

Term	*Definition*
Agency	The European Union Agency for Railways such as established by the Agency Regulation
Agency initiative Output	Tangible products or activities or actions as a result (fully or partially) of an agency initiative, aimed to transfer good practice to other areas and actors that could benefit from the implementation of that good practice.
Agency initiative Outcomes	The likely or achieved short-term and mid-term effects of an initiative outputs.
RailNetEurope	An association set up by a majority of European rail infrastructure managers and allocation bodies to enable fast and easy access to the European rail network. Its aim is to provide support to railway undertakings in their international activities (both for freight and passengers) and to facilitate the simplification, harmonisation and optimisation of international rail processes. The association also carries out a number of projects co-financed by TEN-T funds with a purpose to help with implementation of the rail freight corridors as set out in Regulation (EU) No 913/2010
Scenario	A description of just one way that an actor can accomplish a particular goal. Written as a series of steps performed by actors or by a solution.
Specific goal	A state or condition that an actor is seeking to establish and maintain.
Use case	A description of the observable interaction between an actor (or actors) and a solution that occurs when the actor uses the system to accomplish a specific goal.
Use case diagram	A graphical description using the unified modelling language that captures all actors and use cases involved with a system or product.

1.2.2. Specific definitions business analytics/business architecture

Term	*Definition*
Actor (use case)	Any person or system external to the solution that interacts with the solution. A use case is started by an actor, referred to as the primary actor for that use case. Other actors who participate in the use case in a supporting role are called secondary actors.
Flow of events (use case)	A set of steps performed by the actor and the solution during the execution of the use case. Most use case descriptions separate out a basic, primary, or main success flow that represents the shortest or simplest successful path that accomplishes the goal of the actor. Use cases may also include alternative and exception flows. Alternative flows describe other paths that may be followed to allow the actor to successfully achieve the goal of the use case. Exception flows describe the desired response by the solution when the goal is unachievable and the use case cannot be successfully completed.
Goal (use case)	The goal is a brief description of a successful outcome of the use case from the perspective of the primary actor. This acts as a summary of the use case.
Name (use case)	The use case has a unique name. The name generally includes a verb that describes the action taken by the actor and a noun that describes either what is being done or the target of the action.
Precondition (use case)	A precondition is any fact that must be true before the use case can begin. The precondition is not tested in the use case but acts as a constraint on its execution
Post-condition / guarantee (use case)	A post-condition is any fact that must be true when the use case is complete. The post-conditions must be true for all possible flows through the use case, including both the primary and alternative flows.
Trigger (use case)	A trigger is an event that initiates the flow of events for a use case. The most common trigger is an action taken by the primary actor.

1.2.3. Specific definition for data management

Term	*Definition*
Master data	Data about the business entities (persons, organizations, places, things). They represent the authoritative, most accurate data available. They are held in a single repository following data governance, and linked to by all other systems and reports that need to reference the authoritative source.
Reference data	Data used to characterize or classify other data, or to relate data to information external to an organization (e.g. Country code, validity status)
Data warehouse	A repository of historical data that are organized by subject to support decision making in the organization. IT consists of an integrated decision support database and the related software programs used to collect, cleanse, transform and stored data from a variety of operational and external sources.

1.2.4. Acronyms and abbreviations

Acronym/Abbreviation	*Meaning*
ATO	Automatic Train Operation
CCS	Control Command Signalling
CUI	Common User Interface
db	Database
ECVVR	European Centralised Virtual Vehicle Register

Acronym/Abbreviation	*Meaning*
ERA	European Union Agency for Railways
ERADIS	European Railway Agency Database of Interoperability and Safety
ERATV	European Register of Authorised Types of Vehicles
ESB	Interface Standard Bus
ESG	Economic Steering Group
EVR	European Vehicle Register
IM	Infrastructure Manager
IoT	Internet of Things
KPI	Key Performance Indicator
NPV	Net Present Value
N/R	Not Relevant
NSA	National Safety Authority
OPE TSI	Operation Technical Specification for Interoperability
OSS	One Stop Shop (for vehicle authorization and safety certification)
Pax	Passenger
PRM	Persons of Reduced Mobility
RINF	Register of Infrastructure
RDD	Reference Document Database
RMMS	Rail Market Monitoring Scheme
RNE	Rail Net Europe
RU	Railway Undertaking
SERA	Single European Railway Area
SPD	Single Programming Document (for the Agency)
SRD	Single Rule Database
TAP/TAF	Telematics Application Passengers/Freight
TIS	Train Information System (RailNetEurope)
TF	Task Force
TSR	Temporary Speed Restriction
TTR	Timetable Redesign
UIC	International Union of Railways
VKM	Vehicle Keeper Marking Register

2. Background

2.1. Introduction

Current Agency work in the area of registers and databases aim at responding to business and operational needs of the railway sector, notably deriving from railway undertakings (RUs) along with the related needs of Infrastructure Managers (IMs) and other concerned parties.

To this end, the ERA Single Programming Document (SPD) 2019 has the Objective 4.1 “Collect and analyse evidence for decision making”. Under this Objective, one action is related to this Ad Hoc Task Force.

2.2. Purpose

Alignment of the Agency’s work in the area of registers and databases with the business needs of the railway operators and other concerned stakeholders.

2.3. Objectives

The objective of the ESG Task Force is to collect evidence on **the outcomes** of the Agency current and possible future initiatives on registers and databases directly used when operating trains alongside the evidence on the **final sector impacts** and to quantify the relation between the two.

This is to be done by identifying:

› use cases underpinning the current environment of the railway-related databases and registers;
› additional use cases including their estimated benefits (outcomes), within the railway operational phases which are missing today in the scope of the Agency’s registers and databases.

In addition, the work of the TF will contribute to:

› the high-level description of a proposal for target architectures (“to be” future state) and of the current baseline architecture (“as is”, current state) covering the business, data, application, and technology domains for the Agency’s environment of the railway-related Databases and Registers,
› the definition of a long-term vision for railway for freight and passenger operations

2.4. Scope

The Agency’s registers and databases, as described in the SPD and referred to in the Agency Regulation 2016/796 Article 37 “Registers and their accessibility”, which respond to the business/operational needs of the RUs along with the related needs of IMs and other concerned parties. This implies notably to RINF, ERATV, EVVR, EVR, ERADIS, RDD and OSS.

The TF shall mainly identify the potential of additional use cases within the different operational phases which are today not yet in the scope of the Agency’s registers and databases and which could be considered in future stages within the development of those registers and databases; e.g. currently the infrastructure registers could not include certain use cases linked to last mile operation, sidings or use cases linked to intermodal traffic, e.g. information related to hubs necessary for operators. The interface between RUs and IMs is of a particular interest, including the exchange of data and dedicated files.

The TF shall not investigate the required detailed parameters necessary to fulfil the existing legislative obligations as this task is already covered by the existing workgroups (e.g. user groups of the registers). Similarly, it should not determine parameters needed to accomplish non-mandatory business cases, such as cross-border maintenance planning.

3. Methodology

3.1. General considerations

Agency interventions

Agency interventions should ideally originate from some evidence of a problem and opportunity where there is clear evidence of success because of an action (practice proved to be good), as well as evidence of a lack of such actions in other areas/parts where a potential improvement in performance is envisaged if such good practice is also implemented.

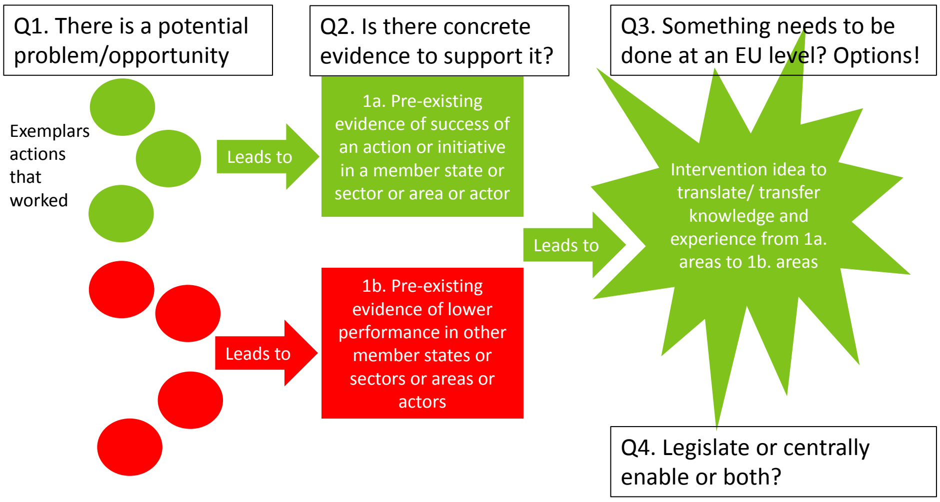

Figure 1: Key questions before defining an agency intervention

Once the need for an intervention is established, then the nature of the intervention needs to be agreed. The type of agency interventions to aid the translation of good practice can be categorised as:

- *Codification in legislative documents and associated guidance, of expected actions aimed at NSAs, competent authorities and other legal entities;*
- *Centralized agency action or output to enable (across member state, sectoral and organizational boundaries) effective sharing of data;*
- *No action by the Agency at all.*

Options for registers/databases

Various roles and functions can be attributed to registers and databases. With the focus on business value of those, the registers/databases contribution, in a generic sense, can be categorized as follows:

A. Data needed to meet legal obligations and needs
B. Case for sharing of data between or getting data from another actor
C. Case for sharing data via a centralised system/register across the EU (centralised management and greater transparency)
D. Case for the specific data configuration
(linkages with other systems to gain more benefits and efficiencies)

Techniques preferences

Since ultimately the optimum target architecture is sought, responding to the business needs of operating RUs, standard business analysis techniques should be applied in the work of the TF. Specifically, the use case/scenario description, logical data model and business architecture techniques will be applied.

3.2. Steps

In order to achieve the objectives defined earlier, the work of the taskforce has been carried out in the following steps:

0. Reiteration of the scope
1. Train service provision phases description
2. Definition of specific business goals related to data occurring in different operational phases
3. Description of business use cases and definition of possible scenarios for use cases
4. Description of target system architecture
5. Identification of specific development needs

Each step requires its own specific methodology and lead to specific outputs (deliverables). A more detailed description of each step follows. Each step can be regarded as a system, with an input consisting of the output from the previous step, process and output (Figure 2). An output of each step provide feedback to the input, i.e. output of the previous step and thus leading to the continuous improvement of the descriptions.

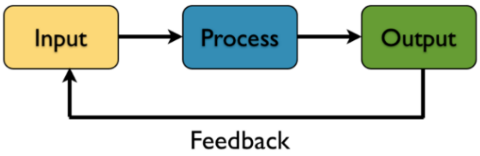

Figure 2: System model for each step

Reiteration of the scope

In a user-centred approach, all solutions should be designed with the user needs and problems in mind. In this task force, the needs of railway undertakings operating trains and thus satisfying the needs of the users (passengers and businesses) are the main focus.

Train service provision phases description

A generic model is sought to describe various phases of train operation, for which specific data needs are expected. A wider view has been taken to cover tasks beyond the own train operation. A high level model of maximum ten steps is meant to provide for a common understanding for any follow-up work; it notably provides a frame for business cases description.

The description used arises from business needs and not from regulatory requirements, although were considered to avoid and potential undesirable conflicts in the follow up work. Here, notably the avoidance of conflicts with OPE TSI is desirable.

Output: A graphical description delivered through this methodological step.

Definition of specific business goals related to data occurring in different operational phases

Single specific business goals are determined, related to data needs, regardless their availability in Agency's registers and databases.

Output: An inventory (list) of goals together with the graphical description.

Description of business use cases and definition of possible scenarios for use cases

A standard description of single business use cases is produced, using a template based on business analytics standard. Possible scenarios are developed on the ways how the underlying business use case is to be accomplished. The following elements are part of the use case description: Name, Goal, Actors, Preconditions, Trigger, Flow of events, Post conditions/Guarantee (refer to Specific definitions for a detailed description). Possible scenario start with “as-is scenario” and develop alternative options satisfying the business needs. Use case sheets are used to provide for case description.

Output: Use cases sheets (including data exchange description)

Description of target system architecture

Logical data model and target system architecture model are determined to best satisfy the most efficient scenarios identified. Alternative proposals are included, if judged relevant.

Output: A series of graphical schemes are to be produced, with necessary accompanying text description.

Identification of specific development needs

Gap analysis is carried out between the current situation and the determined target architecture in order to identify opportunities for actions.

Output: A list of recommendations for further development of Agency’s registers and databases is to be produced. Besides, a recommendations towards the sector will be made.

4. Analysis of RUs data needs

4.1. Reiteration of the scope (STEP 0)

The railway sector is currently undergoing a digitalization revolution, with the use of data increasing exponentially on the background of existing and new use cases. Identifying, describing and analysing all the data interactions would require a substantial effort and time.

As per the Task Force plan, this work should focus on data needs of RUs linked to their train service provision. Although, ultimately, the main beneficiary is the user (passenger or shipper), the information flows between them and the RU is out of the scope for the purpose of this work as illustrated in Figure 3.

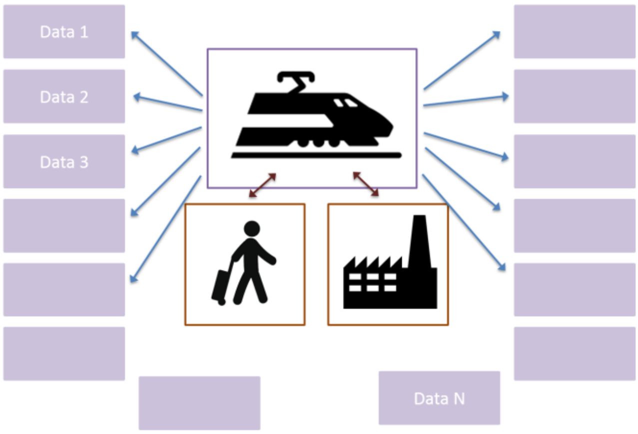

Figure 3: Scope for the business use cases

Both passenger train and freight train service operations are in the scope of the work. It is understood that operation of a passenger train is usually done on a regularly basis, possibly leading to a restricted number of data use cases, once the conditions for the train running has been established.

Using an enterprise architecture model (Figure 4), the scope of the work is limited to business needs, related information and data types. The approach is thus business-driven and not technology (IT systems) driven. The existing EU strategy most visible through the legislative framework, provides a frame for the analysis.

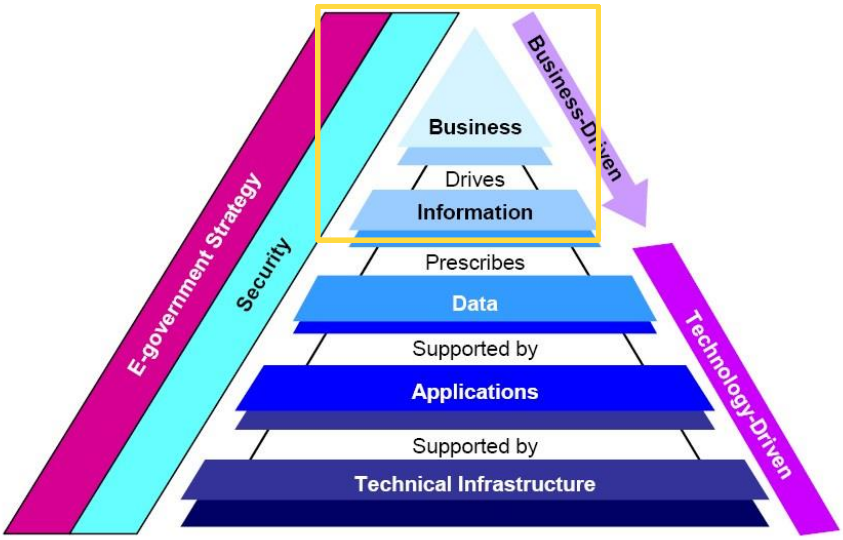

Figure 4: FDIC Enterprise Architecture Framework, based on Zachman framework

All registers and databases, both existing and potential, which (would) answer the data needs of RUs are considered, whereas the following situational criteria are applied:

- *Data is exchanged with another party*
- *Data needs to be kept up to date by another party*
- *There is an obligation or significant sector wide benefit to share data if requested*
- *The data is a snap shot of a state, status and/or description of an asset or operation or service or performance*
- *Data is provided simply but has the functionality to allow the use to focus on critical and relevant data to their operation*
- *Data fairly self-contained in terms of providing answers and there should be clarity around their role in wider decision making*

In practice, registers and databases of data relevant for operation of trains regardless of the data owner or manager, are the focus of the work. Although the end user, or regulatory actors may benefit from the same data as to satisfy their business needs, they are put out of the scope as well.

4.2. Train service provision phases (STEP 1)

The following phases of train service provision have been determined:

Figure 5: Train service provision phases

The phases are common for passenger and freight service operation, however, freight train operation differs from passenger train operations in many aspects (Figure 6), making the first two phases less relevant and the fourth and fifth phase richer, due to higher interaction between the RU and vehicle keeper. In a specific case of a combined transport, the key phase is the train path request and successive phases.

Type of train service	Interaction type	RU offer
Freight	B to B	Product
Passenger	B to C	Service

Figure 6: Aspects of train service operations

The determined phases normally follow on each other chronologically and show dependency. Each phase is described in more detail below:

Phase name	**A - Strategic planning**
Description	Activities to support strategic decision on and planning of a service. Includes market analysis, early assessment of infrastructure, rolling stock assets, human resources, …. In freight, it involves shippers, and is limited to tractive vehicles and un-/loading operations.
Overall goal	To determine the business opportunities by matching requirements and conditions with the RUs assets and strategy.
Specific goals	A1: Identify where (which market) an RU can provide its services, which type of vehicles can be used for potential train services and whether there is available capacity.

	A2: Determine specific conditions and requirements on prospective train operations. A3: Check fleet availability/suitability.
Processes	Strategic planning (e.g. enter or not a market) Specific service planning (e.g. type, characteristics of a service, including routes and their current use/capacity)

Phase name	**B - Vehicle design and authorisation**
Description	Definition of vehicle design characteristics, procurement of vehicles and authorization of the manufactured vehicles. Not relevant in freight, as this is usually assumed independently by a vehicle keeper, and not by a RU. Only applicable for new vehicles, otherwise existing vehicles are used.
Overall goal	To have vehicles ready for use.
Specific goals	B1: Determine vehicle design characteristics for potential train service B2: Determine compatible vehicle types / available vehicles B3: Avoid delays in obtaining a vehicle authorisation (obtain at the lowest costs possible and at shortest time possible)
Processes	Determine vehicle design requirements to ensure that it takes into account of: Infrastructure constraints and possibilities for safe integration with the area of use (network level compatibility), National Rules, Container size, Compliance with TSI in force, etc. Identification of existing vehicle types approved to run on: Networks, areas of use, etc. Identification of NoBos, DeBos to support the verification process (ERADIS)

Phase name	**C - Train path request**
Description	Submitting a path request to IM(s) while respecting and reflecting all requirements. Determining any restrictions on the selected path.
Overall goal	To obtain and maintain access to infrastructure.
Specific goals	C1: Obtain a path matching the requirements. C2: Obtain access restriction information about selected path.
Processes	Iterative Verification of needs and requirements accounting for: Network/line capacity and capability, Start, destination, Route (Tailor made, one-off or pre-arranged path), Framework agreement with IM, If there are dangerous goods and exceptional transport involved in transport, Traction change limitations and issues (diesel, electric, voltage, etc.), Capacity at stations (footfall and trains), Capacity request for facilities (freight train shipment needs), Route compatibility assessment to determine if vehicle will be capable of being compatible with route and operation. Calculation of braking performance and maximal speed permitted.

Phase name	D - Train preparation
Description	Getting a train ready for operation on the path by composing and planning it for own operation.
Overall goal	To have a train ready for running.
Specific goals	D1a: Have adequate vehicles available near to consigner D1b: Assure vehicle configuration of a train matching all requirements. D2: Have the train built-up by shunting services. D3: Have competent and approved staff available. D4: Have a route-book with latest instruction available.
Processes	Verification on vehicles: authorized and registered, compatible with the train path/route and other vehicles Verification of assumptions against actual conditions on the ground notably temporary restrictions related to operational aspects and other needs based on path and vehicle related constraints

Phase name	E - Train operation
Description	The train is ready to operate based on the train movement instructions from IM and runs on the attributed path.
Overall goal	To have a train moving from origin to destination.
Specific goals	E1: Train departs, runs on a planned path and arrives to its destination as planned. E2: When incidents occur during the operation, they are addressed efficiently.
Processes	Information management, in particular between RU and IM Incident management, including contingencies and other measures.

4.3. Specific business goals (STEP 2)

The following specific business goals, which are likely **to be accomplished with a support of data exchange between actors (parties)**, are determined:

Phase	Specific business goal	Pax	Freight
Strategic planning	A1: Identify where (which market) an RU can provide its services and which type of vehicles can be used for potential train services.		N/R*
	A2: Determine specific conditions and requirements on prospective train operations.		N/R
	A3: Check fleet availability/suitability.		N/R
Vehicle design and authorisation	B1: Determine vehicle design characteristics for potential train service		N/R
	B2: Determine compatible vehicle types / available vehicles		N/R

	B3: Avoid delays in obtaining a vehicle authorisation (obtain at the lowest costs possible and at shortest time possible)		N/R
Train path request	C1: Obtain a path matching the requirements. C2: Obtain access restriction information about selected path.		
Train preparation	D1a: Have adequate vehicles available near to consigner (internal) D1b: Assure vehicle configuration of a train matching all requirements. D2: Have the train built-up by shunting services. D3: Have competent and approved staff available. D4: Have a route-book with latest instruction available.		
Train operation	E1: Train departs, runs on a planned path and arrives to its destination as planned. E2: When incidents occur during the operation, they are addressed efficiently.		

*) Not-relevant

4.4. Use cases and scenario (STEP 3)

(Business) use cases linked to the business goals are determined and described using the template defined earlier. Scenarios are developed for their accomplishment. Below is an overview of use cases identified, in which a data exchange between actors occur.

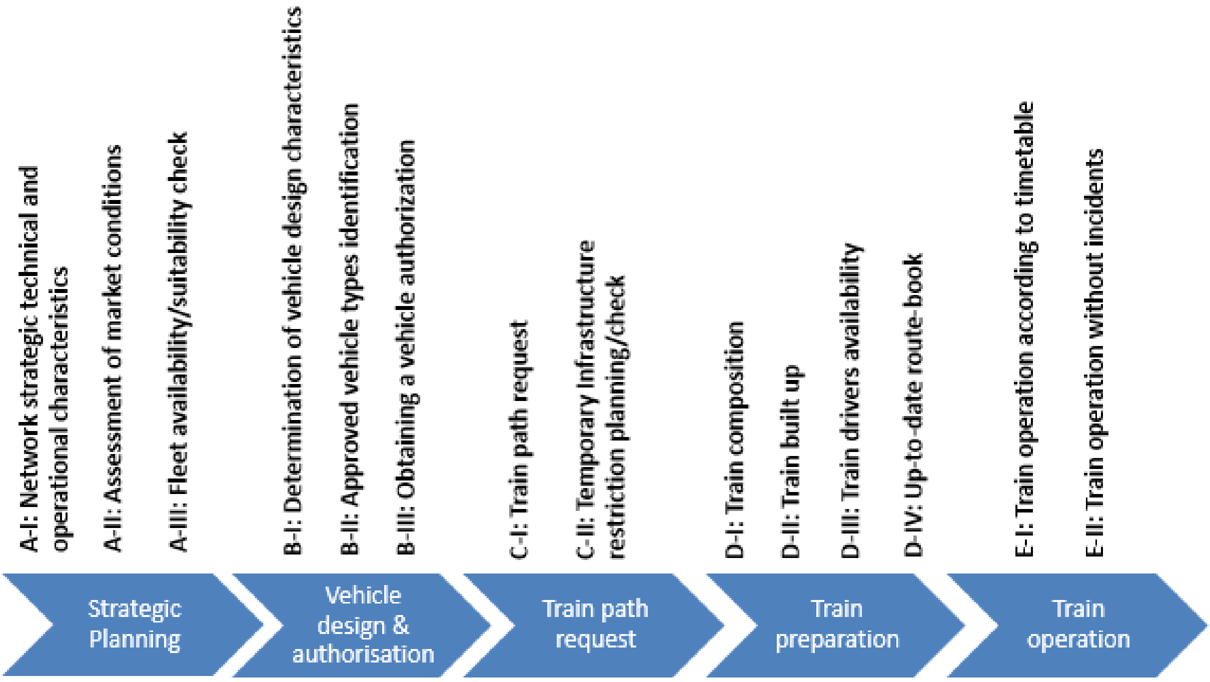

Figure 6: Relevant business use cases related to train service provision phases

As for the development of scenarios for data integration, the following high-level model representation is proposed: The basic level supposes one-to-one ad-hoc interaction between the RU and the data owner, with no prior codification/structure of the data. The intermediate level supposes certain codification of data, including a pre-selection of relevant parameters. The advanced level offers, in addition, a various level of data integration and application functionality.

Degree of data integration

L0 (none)	Reliance on self-analysis supported by data as and when requested from other parties and provided by them in a non-standard format
L1 (basic)	Reliance on self-analysis supported by data as and when requested from other parties and provided by them in a standard format (defined at MS and/or EU level)
L2 (intermediate)	Reliance on published data (kept up to date by all parties) to get data and then perform analysis
L3 (advanced)	Reliance on published data (kept up to date by all parties) and data of relevance to user is available plus analytical functions to leverage data further, and potential links to other pertinent data

These different degrees of data integration can also be depicted in three dimensional scheme (Figure 7). The three dimensions are Access (ease of obtaining needed data), Codification of data (extent to which common taxonomy has been implemented) and Abstraction (correspondence between data and needs).

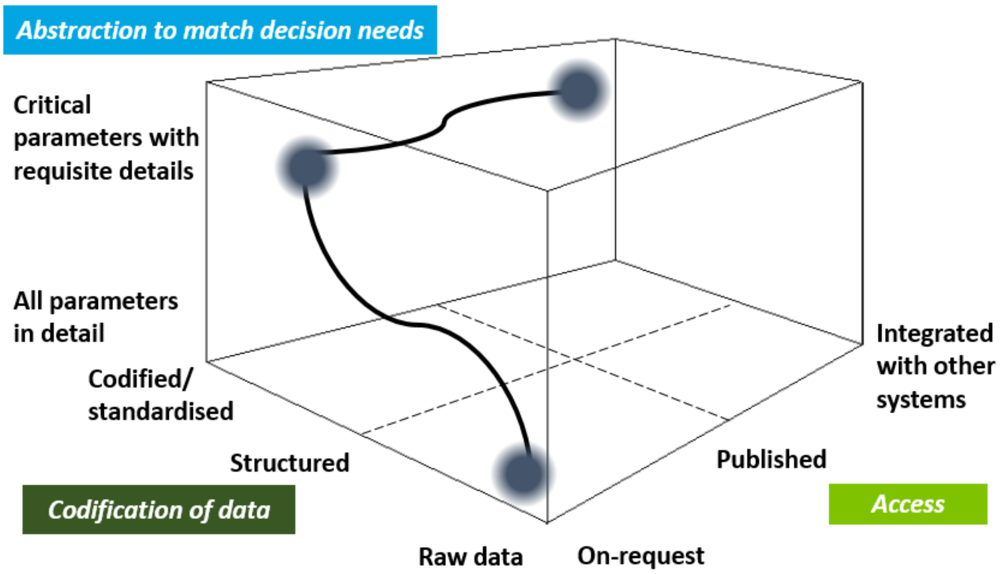

Figure 7: Model for data integration for business use cases scenarios

*Phase A: **Strategic planning***

Use case	**A-I: Identification of technical and operational characteristics of the network**
Content	Technical characteristics / specification of the network (route) Operational characteristics, including capacity and availability. Foreseen development on the network
Goal	A.1 Identify where (which market) an RU can provide its services and which type of vehicles can be used for potential train services.
Actors	RU, IM, Station/Facility owners
Preconditions	None
Trigger	Business opportunity
Flow of events	Own market research > Identification of latest network description > Analysis > Decision
Post conditions	No information gaps exist in understanding technical characteristics of the network.
Scenario 0	RU seek ad-hoc data from different sources (such as Network statements, maps) and perform its own analysis
Scenario I	Technical and operational characteristics from strategic development plans are centrally available, maintained by an entity above IMs. Data on stations and terminals are centrally available. Network statement becomes a structured database. Possibility to enter a line in advance should be more used. Capacity and availability of the network. (Concern the nominal data for vehicle authorisation. In path request, we need the same description of the maintenance works planned or unplanned and their consequence on capacity and availability.)
Scenario II	

Use case	**A-II: Assessment of market conditions**
Content	Existing services on the line, Statistical data, PSO contract timelines Fees and charges
Goal	A2: Identify where RU's business can be performed and which type of vehicles can be used for potential train services.
Actors	RU, IM, Rail market regulator, Ministry of Transport Operators of terminals
Preconditions	None
Trigger	Business opportunity
Flow of events	Finding information on market conditions > Analysis > Decision
Post conditions	No information gaps exist in understanding market conditions
Scenario 0	RUs perform multiple searches of relevant information.

Scenario I	Data on relevant market conditions (already available) are centralized for prospective RUs. Fees and charges lists are centrally available. Fees and charges (CIS). Generators of transport (ports, cities, factories, private yards, etc.).
Scenario II	

Use case	**A-III: Fleet suitability check**
Content	Potential use of already available vehicle (would lead to skipping the next phase). Vehicle-route compatibility assurance.
Goal	A3: Check suitability of already available vehicles.
Actors	RUs, Keepers
Preconditions	None
Trigger	Business opportunity
Flow of events	Finding information on specific conditions > Analysis > Decision
Post conditions	No information gaps exist in understanding specific conditions
Scenario 0	Safety related restrictions currently provided by IMs on ad-hoc basis (not harmonized format, timeliness).
Scenario I	Database of second-hand passenger vehicles (potentially) available for purchase/lease. Data on safety-related restrictions (nominal and permanent data) centralized and made directly available. Possibly part of RINF.
Scenario II	

*Phase B: **Vehicle design and authorisation***

Use case	B-I: Determination of vehicle design characteristics
Content	Design characteristics to ensure that vehicle design takes into account of: Infrastructure constraints and possibilities for safe integration with the area of use (network level compatibility), while keeping a positive route compatibility in mind, National Rules, Container size, etc. **Includes: Vehicle – Network of the area of use technical compatibility assurance.**
Goal	B1: Vehicle design characteristics for potential train service determined.
Actors	RU/Keeper (requester), Vehicle manufacturer, IM, ERA (OSS)
Preconditions	Business decision taken
Trigger	Need for vehicle fleet for planned service
Flow of events	Infrastructure characteristics check > Search for available existing vehicles > Search for available catalogue vehicles > Specification of design requirements > Specification delivery to manufacturer
Post conditions	All information on vehicle design known, so that a resulting vehicle would be granted authorization and be able to run on the designed network.
Scenario 0	Keeper/RU produces technical/functional specifications, reflecting TSI requirements, national rules of area of use OR RU delegates the task to a manufacturer.
Scenario I	Based on RINF and ERATV, additional design characteristics could be defined by the RU.
Scenario II	

Use case	B-II: Approved vehicle types identification
Content	TSIs, national rules, infra-specific requirements. Includes derogations to TSIs. Vehicle-route compatibility assurance.
Goal	B2: Compatible vehicle types / available vehicles determined.
Actors	RU/Keeper (requester), NSA/ERA
Preconditions	Known vehicle characteristics requirements
Trigger	Need for vehicle fleet for planned service
Flow of events	Identification of matching vehicle types > Identification of available vehicles
Post conditions	Vehicle compatible with the designated area of use.
Scenario 0	Ad-hoc search and matching of ERATV and RINF data.
Scenario I	Based on RINF and ERATV data, compatible vehicle types can be determined. [corridors/areas of use recorded for vehicle types in ERATV]
Scenario II	

Use case	B-III: Obtaining a vehicle authorization
Content	Identification of NoBos. Assessment of TSIs (including derogations to), national rules, infra-specific requirements.
Goal	B3: Any delays in obtaining a vehicle authorisation (obtain at the lowest costs possible and at shortest time possible) avoided.
Actors	RU/Keeper/Manufacturer (requester), NSA/ERA
Preconditions	Known vehicle design requirements
Trigger	Need for vehicle fleet for planned service
Flow of events	Request for authorization > Authorization granted
Post conditions	Vehicle authorized and allowed to operate on designated route.
Scenario 0	RU/keeper produces and submits an application file. Iteration with ERA via OSS. Authorization granted (4 months' timeframe).
Scenario I	Integration of ERATV/EVR data into OSS.

*Phase C: **Train path request***

Use case	C-I: Train path request
Content	Path request, granting of a request. Two main options: catalogue train paths, ad-hoc (new) train-paths. Taking/or not into account of the temporary capacity restrictions. **Includes: Envisaged train-intended route technical compatibility assurance.** *(related to B-I compatibility check; depends on country practice, in some cases, the test is organized at this stage to confirm the vehicle compatibility.)* Includes braking capability and maximum speed of the train for the given route (TSI OPE). Providing by the IM to the RUs of the information necessary to calculate the brake performance and the maximum allowed speed (OPE TSI, Annex, section 4.2.2.6.2). TAF/TAP messages on path request management (e.g. Path request, path confirmation)
Goal	C1: Obtain a path matching the requirements.
Actors	RU, IM
Preconditions	Assets (with authorization) for operation ready.
Trigger	
Flow of events	Finding requirements on path request > Path request preparation > Path request submission > Granting of a path > After path allocation, modification of the path by the RU or the IM > Alteration by IM
Post conditions	Access granted.
Scenario 0	Ad-hoc use of catalogue pre-designed train paths. (Exists centrally for corridors (OSS) and sometimes nationally.) The path request is submitted through the path messages of TAF/TAP TSI.
Scenario I	All available train path offer is centralized for all paths and could be requested by the RU using the path messages of TAF/TAP TSI. Includes or not the temporary speed restrictions (TCR).
Scenario II	A Corridor-wide OSS for paths exists, using the path messages of TAF/TAP TSI. Includes or not the TCR (temporary capacity restrictions).

Use case	C-II: Temporary infrastructure restriction planning/check
Content	Data on infrastructure planned restriction
Goal	C2: Obtain access restriction information about the offered capacity or about selected path
Actors	RU, IM
Preconditions	RINF data up to date to feed a special restriction database; Governance in place to deliver real-time data by IMs.
Trigger	Any cause of infrastructure restriction, which makes infrastructure parameters different from published parameters in RINF and Network Statement

Flow of events	Path search request > Path search results delivered with restrictions > Path request submission > Granting of a path with commercial and operational conditions, reflecting the state of infrastructure.
Post conditions	Obtained path compliant with the capacity / capabilities of the infrastructure Accurate ETA/ETI information to customers/RUs/IMs
Scenario 0	National unique approaches, additional files to route book.
Scenario I	Harmonized EU-wide approach, with data available through a dedicated OSS, built onto RINF used as topology reference.
Scenario II	

*Phase D: **Train preparation***

Use case	D-I: Train composition
Content	VEHICLES CHECK: Are the vehicles authorized and registered; Are the vehicle compatible with the train path/route and other vehicles PATH CHECK: Check if assumptions (against actual conditions on the ground notably temporary restrictions) related to operational aspects and other needs based on path and vehicle related constraints are still valid TAF/TAP messages on train composition. Includes : **Vehicle-route technical compatibility assurance** Includes: **Train-route technical compatibility assurance**
Goal	D1: Vehicle configuration of a train matching all requirements.
Actors	RU, IM
Preconditions	All vehicles ready for use.
Trigger	
Flow of events	Pool of vehicles > Compatible vehicles > Train design
Post conditions	Compatible vehicles physically available
Scenario 0	RU relies on RINF for route characteristics. Administrative and design data are derived from different sources by the RU (NVRs-ECVVR/EVR, ERATV, TAF RSRDs, Markings on the wagon). RUs collects data from all available sources. If data is deviant, the RU decides based on own risk management which data will be used. This information is sometimes not identical, e.g. a wagon changes the keeper, RSRD updates the information immediately after the change, same for the NVRs-ECVVR/EVR as this foreseen information is entered in advance with validity dates, the marking on the wagon (VKM) might change months later when the wagon is served next time in a workshop. Manual compatibility check of the train with the route is performed. In freight transport administrative and wagon design data derived from several sources by the RU (NVRs-ECVVR/EVR, ERATV, TAF RSRDs, Markings on wagon).
Scenario I	Possible functional improvements on RINF side, facilitating the task of the RU. For vehicle-related data, one reliable source for every information the RU requires for the checks is determined. Ideally all required freight wagon information come from one central source or sources interlinked with clear status of “master-source” and “slave-source” depending on the data. Administrative master data e.g. for administrative and authorization data NVR/EVR, for design data ERATV and any other data TAF RSRD. Manual compatibility check of the train with the route is performed. In case of freight services, for vehicle-related data, one reliable source of information is available (e.g. TAF TSI RSRDs).

Scenario II	Electronic link of master databases for each data category (e.g. ERATV, EVR, VKM Register, RSRDs etc.) to assure data consistency and quality. Wagon markings might only be used as backup if databases are not available. Automated compatibility check of the train with the route is performed.

Use case	**D-II: Train built up (freight trains)**
Content	Any shunting and other needs are assessed, shunting is requested and performed
Goal	D2: Have the train built-up by shunting services
Actors	RU, IM
Preconditions	All vehicle ready for use. (Train-ready message received by IM).
Trigger	
Flow of events	Compatible and authorized vehicles physically ready > Shunting
Post conditions	Train built-up, ready for departure
Scenario 0	Technical vehicle characteristics (types) related to shunting exists, and are recorded in TAF TSI/RSRDs.
Scenario I	Technical vehicle characteristics (types) related to shunting made centrally available (e.g. via TAF TSI/RSRDs directly linked to ERATV) for shunting entities. Temporary vehicle restrictions are recorded in TAF TSI/RSRDs.
Scenario II	

Use case	**D-III: Train driver staff available**
Content	Staff fulfilling regulatory requirements (license/certificate)
Goal	D3: Have competent and approved staff available
Actors	RU, NSA
Preconditions	Path granted, assets available.
Trigger	Path granted.
Flow of events	Determine HR needs, find staff matching the requirements
Post conditions	Approved staff ready to run a train.
Scenario 0	In case of relying on a train driver employed by other operator, a check on the validity of certificate/license is carried out.
Scenario I	Virtual/centralized register of train drivers (maintained by NSAs/EU register) with services (e.g. drivers authorised on given routes and rolling stock).
Scenario II	Partial and generalised Automatic Train Operation Goal A.4

Use case	D-IV: Up-to-date route-book
Content	Up-to-date route book, temporary infrastructure restrictions, ... Linked to C1:Train path request
Goal	D4: Have a route-book with latest instruction available
Actors	RU, IM
Preconditions	Path granted, timetable available.
Trigger	Path granted, timetable available.
Flow of events	RU compiles a route-book based on information from IMs. Needs to integrate the list of temporary infrastructure restrictions.
Post conditions	
Scenario 0	RU compiles a route-book based on information from IMs. Currently, the information provided is not standardized; will come with new OPE TSI / RINF.
Scenario I	A standard is developed on the file containing the information that the IM provides to the RU to design the route book. Micro-level data become centrally available (RINF+). Operation points description (stations/terminals) needed. Schema of the stations/terminals available; pdf schematic picture could be used in a first approach for the schematic details of a station.)
Scenario II	A standard is developed on route-book for all RUs > digitalized solutions are developed. This standard should be based on the granularity needed for ATO, as numerous RUs and IMs currently work on ATO and the relevant description of infrastructure.

*Phase E: **Train operation***

Use case	E-I: Train operation according to timetable
Content	Train runs according to the timetable, without unplanned restrictions.
Goal	E1: Train departs, runs on a planned path and arrives to its destination as planned.
Actors	RU, IM
Preconditions	Assets for operation ready.
Trigger	
Flow of events	Finding requirements on path request > Path request preparation > Path request submission > Granting of a path
Post conditions	Train de-registered from operational systems (taken out of service).
Scenario 0	TAF/TAP messages + TIS[1] data only used during the train operation.
Scenario I	Centralized database of train running messages data to support service quality performance management (TAF/TAP messages + TIS[2]). Centralized detailed data on train running (journey data), needed by successive or substitute carrier, whereas TAF/TAP messages and TIS DB could be used for this purpose.

Use case	E-II: Train operation with/without incidents
Content	Monitoring and contingencies in case of incidents. Operation to required quality standards. Relates to both infrastructure and the train/vehicles. Common performance measurement arrangements, such as KPIs for RFCs.
Goal	E2: When incidents occur during the operation, they are managed efficiently.
Actors	RU, IM
Preconditions	Train runs.
Trigger	Failure/Incident during train operation.
Flow of events	Identification of a failure/incidents > risk analysis > identification of critical vehicles/procedures > preventive actions If incident occurs in operation > efficient management to reduce the consequences and re-establish the nominal operation
Post conditions	Measures identified to prevent recurring of failure/incident. Measures identified to manage efficiently the failure/incident. Train de-registered from operational systems (taken out of service)

[1] Currently available only for internationally operated trains.

[2] Currently available only for internationally operated trains.

Scenario 0	RU uses its own intelligence / Nationally available intelligence / EU-wide intelligence
Scenario I	Occurrence reporting standards and central solutions, whereas TAF/TAP messages and TIS DB could be used for this purpose. Vehicle performance made centrally available. Sharing of experiences of IMs (traffic managers) and RUs.
Scenario II	Efficient Train Supervision with proposals of scenarios to support the traffic management (target: Automatised Train Supervision).

4.5. Data elements for use cases

Following the identification of opportunities (Scenario I/II of each use case) for centrally available data (facilitating the RU business), underlying information and data elements are identified.

Use case	Information	Data elements
A-I: Identification of technical and operational characteristics of the network	Network characteristics	Network technical and operational general constraints such as Class B CCS systems
A-II: Assessment of market conditions	Market	Network statement, Access/operation costs, Contracts in place
A-III: Fleet suitability check	Network compatible vehicle types, Available vehicles	Available vehicles, vehicle types
B-I: Determination of vehicle design characteristics	Vehicle characteristics relevant to the specific networks	Network specific vehicle technical and operational rules and constraints
B-II: Approved vehicle types identification	Network/route vehicle compatibility, approved vehicle types	Technical and operational conditions of use/constraints, approved vehicle types
B-III: Obtaining a vehicle authorization	Status of authorization application	Stages of authorisation, Actions required before progressing to next stage, Application status in relation to stages
C-I: Train path request	Train path catalogue, specific train path	Train path
C-II: Temporary infrastructure restriction planning/check	Temporary infrastructure restriction	Transient infrastructure (incl. planned maintenance)
D-I: Train composition	Route book, Transient technical, maintenance/ performance info	Route book, transient vehicle, maintenance/performance vehicle
D-II: Train built up	Shunting and other restrictions	Shunting and other restrictions
D-III: Train driver staff available	Available approved drivers	Approved drivers, Available drivers
D-IV: Up-to-date route-book	Route book / Data for ATO	Route book / Data for ATO
E-I: Train operation according to timetable	Train running information	Performance records
E-II: Train operation with/without incidents	Train's operational status and operation mode	Safety occurrences and other incidents that can impact train operation

4.6. Function data model

A function model or functional model is a structured representation of the functions (here specific information exchange needed to satisfy the use case) within the modelled system (here provision of train services). A function model is drawn for data elements underlying the information needs identified for single use cases identified earlier (Figure 8). Specific data (bases) are then identified and the relationships depicted with the use of lines and arrows.

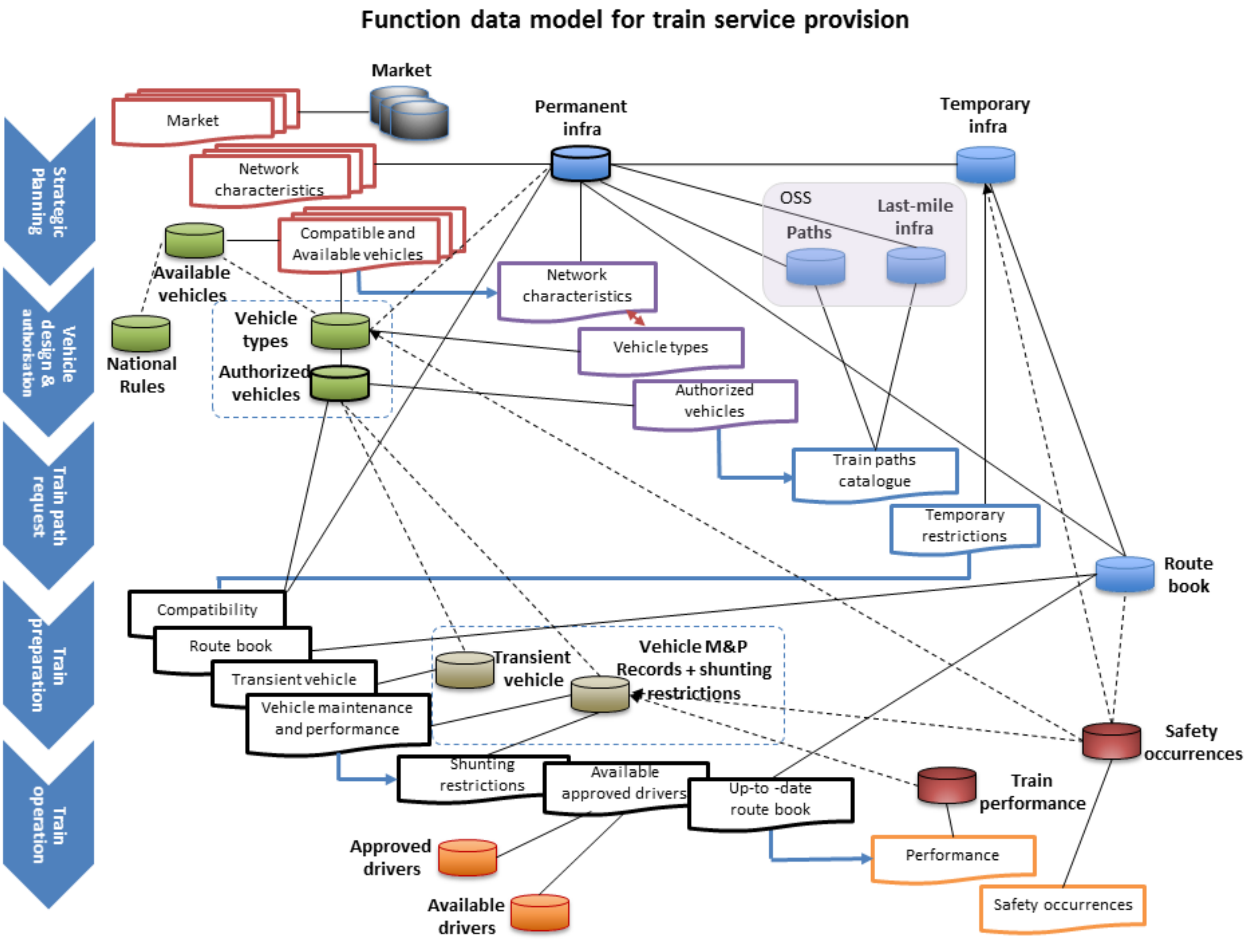

Figure 8: Function model for data elements underlying train service provision

In phase A, strategic planning, all data relevant for assessing business opportunity are grouped under Market database. Network characteristics of interests are those stored as permanent infrastructure data. Types of vehicles compatible with the network are available in a database of vehicle types. While the types of vehicles potentially available for lease/sale (now and in the near future) are available in a database of available vehicle types.

In phase B, vehicle design and authorization, vehicle design characteristics could be determined for the given area of use/network. The compatible vehicles, authorized for use in specific area of use, are registered in a database of authorized vehicles. Route compatibility check could contribute to the vehicle design.

In phase C, train path request, a request can be satisfied thanks to two databases: one on description of network/infrastructures and available capacity and one on available last-mile infrastructure. These are grouped under a one-stop-shop, delivering the requested path, tailor-made or pre-arranged. Further, the data on temporary infrastructure restrictions are needed as part of service planning.

In phase D, train preparation, vehicle-route and train-route compatibility must be first established, relying on both vehicle and infrastructure data. In both cases, both permanent and temporary/transient data need to be used. The temporary vehicle data consists of transient technical data, maintenance and performance records and shunting restrictions. Route book data are needed for train composition and for service running. To assure availability of a train driver, a database of available drivers can be consulted if a “drivers market” is set up. To check the validity of a licence/certificate of a given driver, a database of approved drivers can be consulted.

In phase E, train operation, train performance data are collected, using TAF/TAP messages, for accounting and statistical purposes. Safety occurrences are recorded and stored in a database of safety occurrences, relying on several other databases for the details on vehicles, infrastructure and staff involved.

The function data model makes appear the single most important data elements, which are “permanent” infrastructure and vehicle data. They are most referenced and other data elements and thus constitute the ideal reference data. Similarly, safety occurrence and route book data elements have a high degree of data integration, but do not constitute reference data in the traditional sense.

As regards “temporary” data, operational data on vehicle, which stretches from transient technical data through technical and operational data to temporary restrictions and performance records (such as mileage and disturbances in operation), form a single most important operational data package. (Note, that for passenger train service operation, it would be much reduced, in the content, compared to freight train services.)

5. Data architecture (STEP 4)

Data architecture has been consistently identified by CXOs (Chief Experience Officers) as a top challenge to preparing for digitizing business. The data architecture is about "using data effectively and built on a foundation of business requirements[3]". In practice, the data architecture is composed of models, policies, rules or standards that govern which data is collected, and how it is stored, arranged, integrated, and put to use in data systems and in organizations.

5.1. Data flow design

A data flow model depicts how the data moves through business processes and systems. Two representations are commonly used: a two-dimensional matrix and flow diagram. Hereby, a matrix is developed, giving a clear overview of what data the processes create and use. The group business processes are here equal to phases of train service provision and entities are larger blocs of data elements identified previously (Figure 9).

BUSINESS PROCESS GROUPS

MAJOR ENTITIES	Strategic Planning	Vehicle design & authorisation	Train path request	Train preparation	Train operation
Market	○		○		
Available vehicles	○	○		○	
Vehicle types	○				
Infrastructure permanent	○	○	○		
Facilities	○		○	○	
Paths			○		
Route book					○
Infrastructure transient				○	○
Drivers	○				○
Vehicle operational				○	●/○
Train performance					●/○
Occurrences					●/○

● CREATE ○ READ/USE

Figure 9: Data flow depicted in the matrix

(Note that the matrix presented contain data workflow relevant for RUs only and similar, complementary matrices could be developed to cover IM, RU, Keeper and NSA needs, as to identify possible synergies.)

The data flow matrix shows that:

- First, RUs are mainly consumers of data created by other entities under their business process groups (very often regulatory requirements).

- Second, the use of data typically occur repeatedly under multiple processes groups (phases), bringing about certain complexity in requirements on the data.

- Third, in the train operation group business process groups (phase), data are both created and consumed, suggesting a need for a more detailed (disaggregated) description of the processes.

---

[3] McKinsey, 2017: Why you need a digital data architecture to build a sustainable digital business, [Online article]

The multiple use of the same data (in different phases) and the fact that they typically originate from organizations external to RUs means that a detailed understanding of use cases under business processes use cases is prerequisite for designing the suitable taxonomy for data entities and data governance structure.

A further dive into specific business process groups unveils major data entities (elements) needed for specific use cases. Given its importance for stakeholders (as per RU survey), a specific data flow matrix is developed for the vehicle/train – area of use/route compatibility assurance use cases. These occur under different phases of train service provision, sometimes repeatedly (Figure 10).

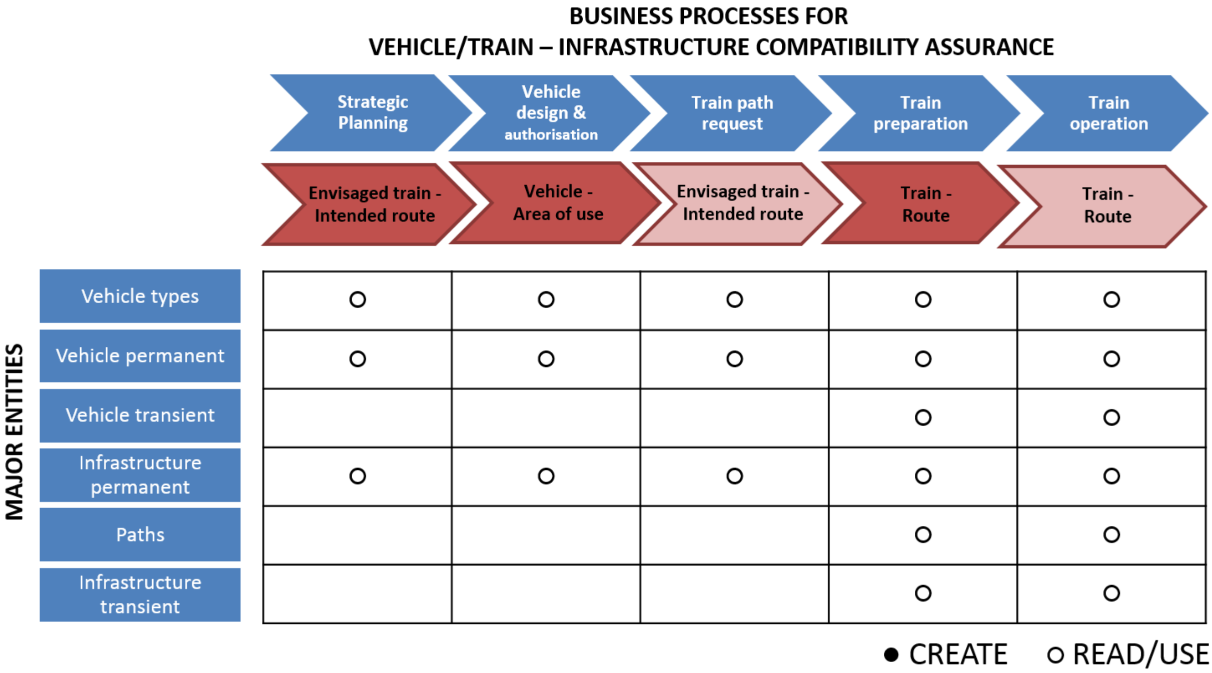

Figure 10: Data flow for vehicle/train – route compatibility assurance depicted in the matrix

The matrix shows that train-route compatibility assurance is more complex than other use cases (under compatibility assurance), as it relies on six data entities (compared to three in other cases). It is notable that train-route compatibility is an operational task relying on operational data and as such might be best assured by the operational sector itself (see conceptual model in the next chapter).

5.2. Conceptual data model

A data model organizes data elements and standardizes how the data elements relate to one another. In this work, we focus on conceptual models, which aim is to establish the entities, their attributes, and their relationships. Unlike other types of data models, the conceptual data model offers organisation-wide coverage of the business concepts. It is designed and developed for a business audience. It is developed independently of hardware specifications like data storage capacity, location or software. The focus is to represent data as a user will see it in the "real world." Since the scope of the work of the TF is limited to high-level architecture, the single attributes (in a data model) are not described and analysed in detail.

A high-level elementary data model is proposed to depict similarities and relationships between different data elements, which were identified previously. We first categorise the identified data per subject: market, infrastructure, vehicle and train. We then attempt to split the data per its nature: regulatory/administrative data and operational/business data. The data subject categorization is motivated by the logic of data origin and data owners/providers, whereas the nature of the data categorization is meant to reflect motivational logic and consequently the likely ideal steward of the data (Figure 11). It could thus be considered as an idealistic data model for the identified use cases.

In principle, market/administrative data elements are needed to either address inefficiency in the market and/or to enhance safety oversight, while the operational/business data (made centrally available) are motivated by business interests.

Market administrative data elements are those contained in the network statement, other data on the given railway market and data on approved train drivers. Market operational data elements are notably those on (potentially) available train drivers.

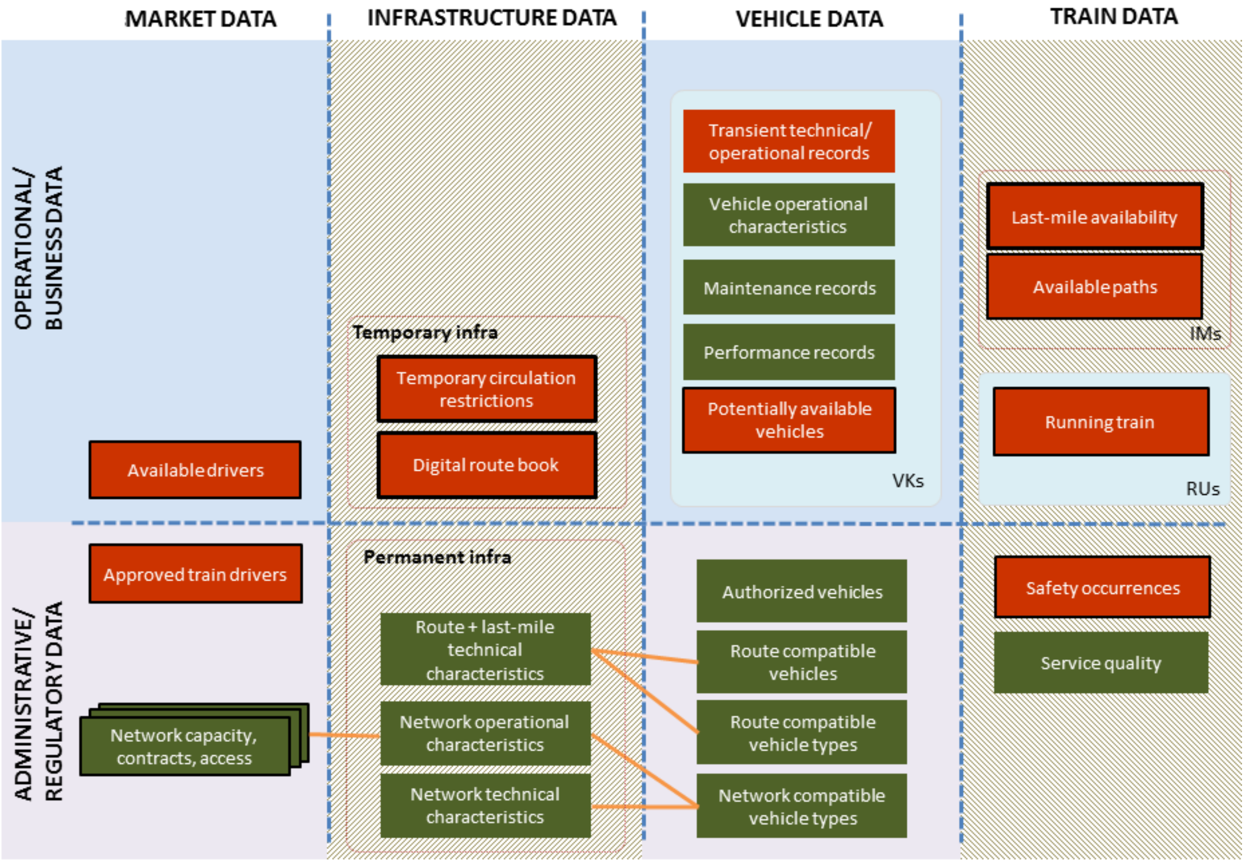

Figure 11: Elementary model for data elements

Infrastructure administrative data elements are those providing the overview of the potential of the given network for train services and those related to design and authorization of vehicles. They can be viewed as "permanent" due to their relatively long validity, extending beyond the validity of a regular (yearly) timetable. Infrastructure operational data includes data for vehicle- and train- route compatibility check, digital route book, safety and temporary restriction and the last mile data. They could be viewed as "temporary" as they are subject to changes over time within the validity of a regular (annual) timetable.

Vehicle administrative data elements and data on authorized vehicles facilitating the safety oversight and vehicle technical/operational characteristics facilitating a route-vehicle compatibility check, meant to help overcoming potential market barriers.

Administrative/regulatory train data are data on service quality and performance (volumes). Operational/business data elements are those needed and exchanged in single train operation.

In the elementary model for data elements (Figure 11), the red box colouring is used to identify those data elements that are nowadays not available centrally. Some elements are grouped as to highlight either their common nature/treatment or origin/ownership.

The conceptual data model provides for rationalized grouping of certain data elements. It notably suggests that half of them could be regarded as administrative/regulatory data. It further shows that all new data elements fall under operational/business data meaning that the newly identified use cases should be left to the sector to implement. For the two specific improvement use cases under administrative/regulatory data, the data are available at country level, with partially compatible taxonomies, so the improvement would require a minor effort in data compatibility assurance and in setting up the communication among existing IT systems/databases.

5.3. Fitness check

A critical review is carried out to determine whether data elements in the elementary data model above are nowadays centrally available and whether their nature/treatment is in line with the logical categorization proposed.

Market data

Administrative/regulatory data are fragmented, scattered and not available centrally for RUs. Specifically, network statements may be stored centrally, but only as text files, and not as a database. Market data in RMMS are available to MSs for consultation, and only a selection of data is published, with important delay, in statutory RMMS reports, by the Commission. Some data, such as contract duration information may be hard to find at all.

Data on approved train drivers are available at the national level, what corresponds to the current typical modus-operandi in international train service provision. However, for an RU, wishing to organize a cross-border service without switching a train driver and making use of another operator for operating the journey log in other country, such a database would be beneficial. (RU established in any Member States may nowadays consult the rail driver register of any other Member States, but this solution is less efficient than a centralized database).

Infrastructure data

RINF has been introduced to include all relevant administrative/regulatory data necessary in business planning (including early vehicle-route compatibility check), where the access to data for identified use cases was viewed as inefficient. Last-mile infrastructure access data are nowadays understood as operational data left out from regulatory reporting. This may not be an optimal situation given reported inefficiencies and access/use difficulties reported by some RUs. Single Rule Database stores data on specific operational conditions on the network and thus suits the purpose and the model.

Therefore the top priority may be to improve the description of infrastructure in RINF to reach the correct granularity for the use cases. It is notably proposed to focus as target infrastructure description on the granularity needed for ATO to improve RINF, as numerous RUs and IMs currently work on ATO. This granularity is compliant with the infrastructure related use cases (B-II, C-I, C-II, D-I, D-IV, E-II, etc.). A study on this should be achieved in the frame of RINF and CCS (ATO).

RINF is foreseen to host more detailed data needed for train operation (e.g. route book, or even ATO-required data), or temporary circulation restriction. These data may be better shared via an (external) user-oriented database, as to gain in efficiency.

Vehicle data

ERA TV and ECVVR contain administrative/regulatory data on vehicle types and vehicles, yet some hindrance prevails in their (public) availability. Route-compatible vehicles/types are nowadays not available and this not even when making a combined use of several registers.

Centrally available operational/business data elements exist for freight. The data are maintained by keepers. In case of passenger services, data on potentially available vehicles are not centrally available, although this report has identified a business case for it.

Train data

Service quality reports are available in ERADIS, but not as a database. They contain information harmonized to a certain extent, but practically no structured quantitative data. Aggregated safety performance data are available in ERAIL, alongside the records on investigated accidents. Both are of limited used to RUs.

Running train data such as locations, sections worked out, type of services, specific goods transport (such as dangerous goods) and available through TAP/TAF, yet not made centrally available (although some exists for TIS for train running). As some of those may constitute regulatory (statistical) data reportable under RMMS, there may be a certain case for their introduction. The last mile service availability information is of a business interest, and would be useful if available centrally.

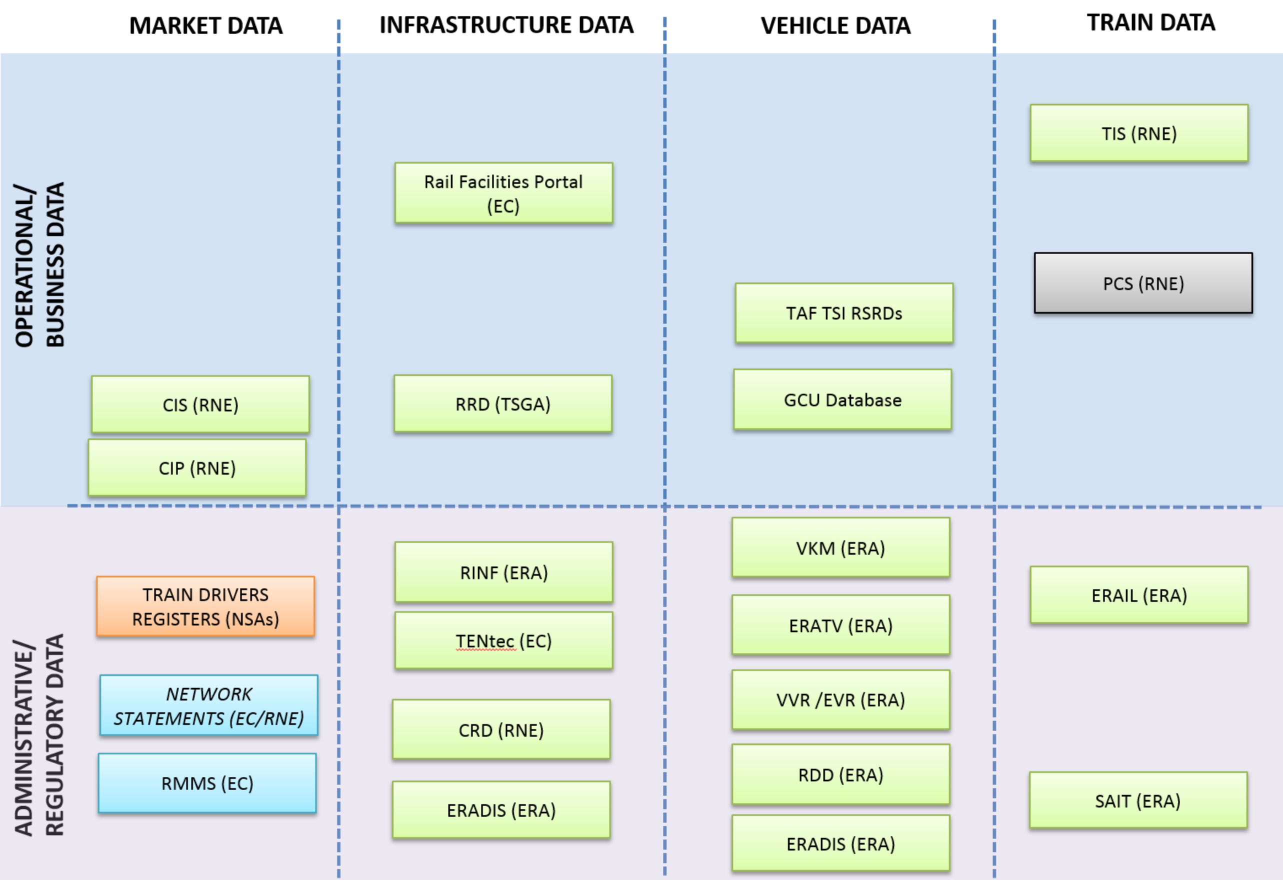

Figure 12: Existing central databases (RU use cases)

Train paths data are stored by single IMs and only for RFC paths, made available centrally by RNE. (In practice, the RNE is connecting information on path details provided by IMs to create an international train path. Later, IMs provide train running information and RNE can connect on behalf of IMs the national information to an international one.) There is no rational reason for not making all types of paths on the entire SERA network centrally available.

Last-mile infrastructure availability for train handling, specifically opening hours, is not centrally available (not even in the last mile portal).

The assessment was carried out against the existing centralised databases providing data to satisfy the use cases identified earlier. The databases identified are listed in Figure 12.

Altogether nine data elements appear to be a potential candidate for a new centrally available database, either alone standing, integrated into existing database/register, or virtually connected database network. An economic assessment is undertaken in chapter 6 to determine their relevance for fulfilling the use cases and delivering added value to RUs.

In parallel, for four data elements, a case exists for enhancement of data availability through a central database. In case of approved train drivers, market access data, safety occurrences and available paths, the harmonization and centralization effort would be required to enhance the value of data (made them centrally available).

A high-level comparative analysis of data models used in existing central registers/databases hosted by ERA, allow for the identification of data entities used in several registers/databases. Furthermore, it gives emergence of the needs and priorities in railway master data management.

Register/ Database	**RINF**	**EVR**	**ERATV**	**VKM**	**RDD**	**CCM**	**ERADIS**				**ERAIL**	**NOTIF-IT**	**Count**
							SC	**L**	**ECM**	**NoBo**	**INV**		
Taxonomy source / Major entities	2014/880/EU	(EU) 2018/1614	2011/665/EU	Application guide	2009/965/EC	CCM Procedure			User manual (ERA)		Tutorial (ERA)	User manual (ERA)	
Organization	x	x	x	x		x	x	x	x	x	x	x	11
Location	x	x	x	x	x		x	x	x	x	x	x	11
Market		x	x		x	x	x	x					6
Validity status	x					x	x	x		x		x	6
Infrastructure	x	x				x					x		4
Vehicle		x	x		x						x		4
Train						x					x		2

The organization and location are the entities with the most frequent occurrence, used in practically all registers. They are thus perfect candidates for master data management. Other priorities include market data, where additional corresponding entities exist in databases managed by EC, validity status, infrastructure and vehicle entities.

More specifically, these build a case for reference and master data management for the following data entities and objects:

Reference data	**Master (primary) data**
Organization types	Organization
Location (GIS, country codes)	Location (addresses)
Certificate/authorization/license validity	Certificate/authorization/license
Train paths	Train ID
Vehicle types	

While for reference data, only a few examples are showed, the list for master data is likely more complete. A major effort may be needed to map the data models in all existing databases mentioned above, as to identify inconsistencies in reference data use. An example of the results of this effort would be an inconsistent use of country codes: ISO 3166 alpha-2 codes for all countries except for Greece and the United Kingdom codes for country of registration and alphabetical and numerical code based on UN convention on road traffic for coding of countries in vehicle markings in ERA EVR. Own numerical list of countries in TAP-TAF; or an inconsistent use of organization ID: ERA EVR (registered business number) versus TAP-TAF (VAT code).

As regards the master data, while for legal master data (certificates/authorization/license), a single master database has been established within an EU framework, the party master data (organization) and location master data are currently not managed in a proper way. The Agency has already started to work on the organizational code list and on location data for its stakeholders (in the SRM tool), but this effort has not yet been completed.

6. Economic assessment

6.1. Improvement options

Improvement options refer to increase in data integration for single data elements, whereas the four different levels of data integration defined earlier are considered:

Data integration level	Description
L0 (none)	Reliance on self-analysis supported by data as and when requested from other parties and provided by them in a non-standard format.
L1 (basic)	Reliance on self-analysis supported by data as and when requested from other parties and provided by them in a standard format (defined at MS and/or EU level).
L2 (intermediate)	Reliance on published data (kept up to date by all parties) to get data and then perform analysis.
L3 (advanced)	Reliance on published data (kept up to date by all parties) and data of relevance to user is available plus analytical functions to leverage data further, and potential links to other pertinent data.

For each data elements (corresponding to its use case), the current level of data integration is determined. In the following step, the desired level of data integration, corresponding to use case scenario 1 (or 2) is provided.

Use case	Data elements	Current level of data integration (Scenario 0)	Desired level of data integration (Scenario 1/2)
A-I: Identification of technical and operational characteristics of the network	Network technical and operational general constraints such as Class B CCS systems	L2 RINF technical L1 Last mile infra L0 Operational	L2/3 L2/3 L2/3
A-II: Assessment of market conditions	Network statement, Access/operation costs, Contracts in place	L1 Network capacity L0 Fees/Charges/ Contracts/Access	L2/L3 L2/L3
A-III: Fleet suitability check	Available vehicles, vehicle types	L0 Available vehicles L2 Vehicle characteristics in registers (ERATV, NVR-ECVVR-EVR) L1 Freight vehicle characteristics in vehicle dbs (RSRD)	L2/3 (market-driven) L3
B-I: Determination of vehicle design characteristics	Network specific vehicle technical and operational rules and constraints	L2 Network technical (in RINF)	L3
B-II: Approved vehicle types identification	Technical and operational conditions of use/constraints, approved vehicle types	L0/2 Vehicle design characteristics (ERATV/EVR) - (limited usefulness due to absence of data and linkages)	L3 – Need to be coded for area of use

		L2 Network data (as in RINF)	L3
B-III: Obtaining a vehicle authorization	Stages of authorisation, Actions required before progressing to next stage, Application status in relation to stages	L0	L2 (need for data integration ERATV / EVR and OSS)
C-I: Train path request	Train path available	L2 / L1 (for international paths only, hosted by RNE / for national, TAF/TAP formats)	L3
C-II: Temporary infrastructure restriction and TCR planning/check	Transient infrastructure (incl. planned maintenance)	L0 [Pilot under evaluation at RNE]	L3
D-I: Train composition	Route book, transient vehicle, transient infrastructure, maintenance/performance vehicle	L0 Route book L0 Transient vehicle data L0 Transient infra L0 Maintenance/ perf	L3 L3 L3 L3
D-II: Train built up	Shunting and other restrictions	L0 (possibly stg in RSRD)	L2/3
D-III: Train driver staff available	Approved drivers, Available drivers	L0	L2
D-IV: Up-to-date route-book	Route book	L0	L2/3
E-I: Train operation according to timetable	Performance records	L3 (RNE members for international services + new provisions in TAF/TAP for IM storage) L1 (national)	L3
E-II: Train operation with/without incidents	Safety occurrences and other incidents that can impact train operation	L0 L2 (ERA SAIT Tool)	L3

6.2. Qualitative assessment of improvement options

Qualitative assessment (expert opinion) was carried out for the data elements and their suggested improvement options, whereas per se, the maximum improvement options corresponding to advanced data integration (level 3), are assessed.

Relative costs and benefits of improvement cases were identified for single data category and assessed, using a Delphi technique[4] with two rounds of expert opinion survey. In the first round, TF members provided individual assessment of first followed by a group discussion on aggregated result, followed by a second individual assessment leading to final aggregated results.

TF members were asked to subjectively assess relative costs and benefits of 21 improvement cases identified for single data category, using the four-scale: none/minimal=*, low/modest=**, Moderate =***, High=****, where:

'High' and 'Minimal' benefits *respectively designate a potential to* ***significantly (for High) or have no noticeable (for minimal) improvement in the effectiveness and/or efficiency*** *for railway undertakings and other relevant actors* ***when taking decisions associated with the Train Service Provision Phase the data category relates to.*** **The maximum improvement options correspond to moving to the advanced data integration level (level 3).**

And

'High' and 'Minimal' costs *respectively designate a potential to* ***significantly (for High) or have no noticeable (for minimal)*** *increase in the* ***effort and cost associated with any or all of the following aspects:***

- *Development and implementation of the system and the associated data architecture*
- *Provision and/or collection of accurate and timely data*
- *Changes/redesign of existing systems, data structures and formats including the distributed systems that would input/interact/integrate with the central system*
- *Ongoing maintenance, checking and assurance so that the data being shared and already stored remains fit for purpose*
- *Access to and dissemination of the data outputs with the users*

Altogether nine TF members took part in the exercise (no ERA staff members took part). The first round of the assessment led to a relatively homogenous results. Among 21 improvement cases, only 7 had B/C ratio higher than one, while nine had a B/C ratio close to 1 [0.9; 1.1]. The second round of the assessment resulted in more distinctive results, showed in Figure 13.

[44] The Delphi technique (also referred to as Delphi procedure or process), is a method of congregating expert opinion through a series of iterative questionnaires, with a goal of coming to a group consensus.

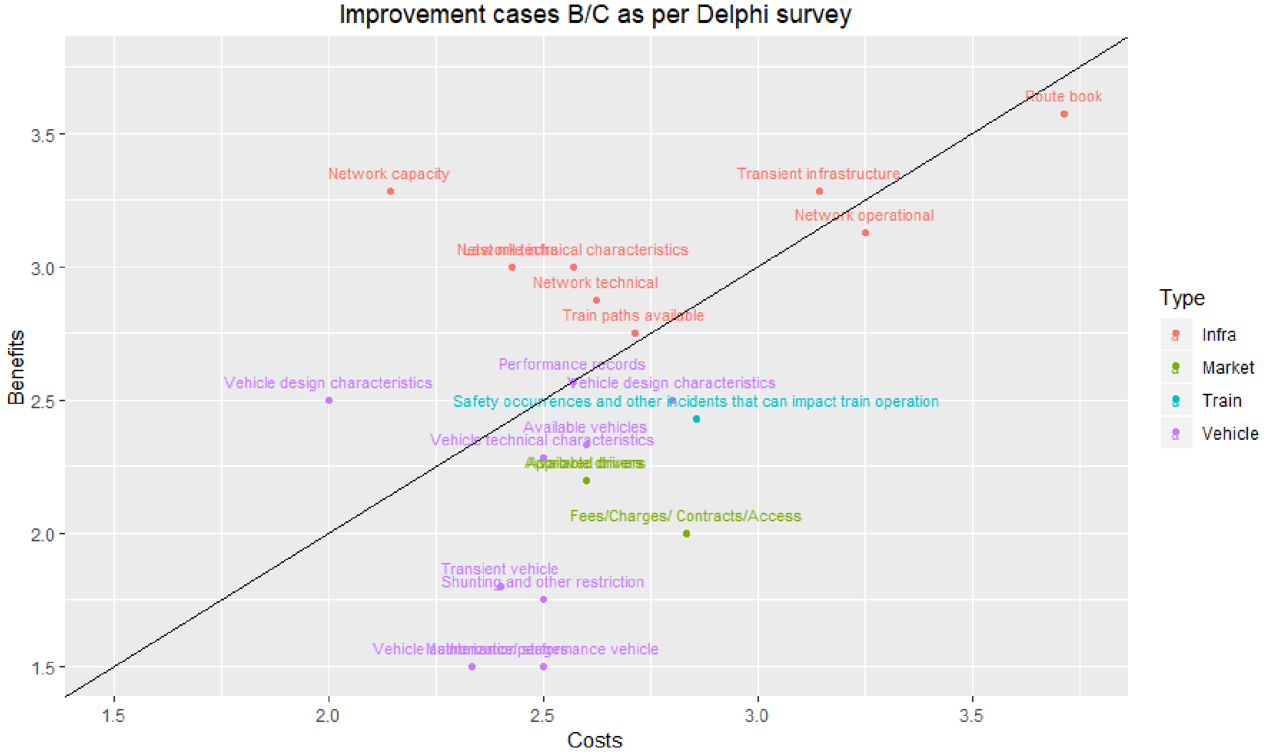

Figure 13: Expert group (Delphi) attributed costs and benefits of improvement cases

The same results are summarized also in tabular form, sorted descending by B/C ratio. This allows identification of more and less cost-efficient improvements. It should be noted that the B/C ratio here reflects the relative perception of benefits vs. costs (rather than quantified or monetary factors), for example a B/C ratio of 1.1 might provide a far greater benefit than a B/C ratio of 2, therefore should be treated with caution and is only there to aid discussion and debate.

Improvement case	Costs	Benefits	B/C
Network capacity L2>L3	2.14	3.29	1.53
Vehicle design characteristics L2>L3	2.00	2.50	1.25
Last mile infra L1>L3	2.43	3.00	1.24
Network technical characteristics L2>L3	2.57	3.00	1.17
Network technical L2>L3	2.63	2.88	1.10
Transient infrastructure L0>L3	3.14	3.29	1.05
Train paths available L1>L3	2.71	2.75	1.01
Performance records L1>L3	2.57	2.57	1.00
Network operational L0>L3	3.25	3.13	0.96
Route book L0>L3	3.71	3.57	0.96
Vehicle technical characteristics L2>L3	2.50	2.29	0.91
Available vehicles L0>L3	2.60	2.33	0.90
Vehicle design characteristics L1>L3	2.80	2.50	0.89
Safety occurrences and other incidents that can impact train operation 0>L3	2.86	2.43	0.85
Approved drivers L0>L3	2.60	2.20	0.85
Available drivers	2.60	2.20	0.85
Transient vehicle L1>L3	2.40	1.80	0.75
Fees/Charges/ Contracts/Access L0>L3	2.83	2.00	0.71
Shunting and other restriction L0 >L3	2.50	1.75	0.70
Vehicle authorization stages L0>L2	2.33	1.50	0.64
Maintenance/performance vehicle L0>L3	2.50	1.50	0.60

The three cases where the perception of benefits was high are the route book, network capacity and transient infrastructure, with average ratings higher than 3.28. The infrastructure data related improvement cases all have an average benefit rating higher than 3, while all of them, except for the network operational and the route-book (which narrowly missed out) outperformed the perception of costs. Among other improvement cases, only the vehicle design characteristics and (train) performance records have a B/C>1, i.e. the perception of benefits was higher than the perception of costs.

The results suggest that, in general, implementation of improvement cases related to infrastructure data should be privileged over other improvement cases. Three improvement cases are judged to deliver only low/modest benefits (Shunting and other restrictions, Vehicle authorization stages, Vehicle maintenance/performance).

This simplistic assessment that relies on the expert group opinion is a proxy for the results of a systematic detailed CBA delivered for each improvement case. It allows to build a basic hierarchy of the most efficient improvements (in terms of data integration enhancement).

6.3. Quantitative assessment of improvement options

A quantitative assessment is carried out for those improvement options that were identified as most desirable by a pool of RUs involved in this work and which, at the same time, might benefit from Agency’s work (facilitator/enable for taxonomy development, master data definitions, ...).

Model for CBA estimation

Benefits estimation

The measurement/quantification method involves creating a benefits realization chain with three distinct stages:

- Stage 1: Immediate benefits due to the availability of relevant information via central register which would lead to follow-on actions from key actors
- Stage 2: Impact of follow-on actions taken by actors being targeted due to effective immediate impact
- Stage 3: Eventual impact on the sector’s performance due to effective follow-on actions taken by relevant actors

Figure 14 illustrates the specific three stages to link the existence of a register with the actual benefits to the sector.

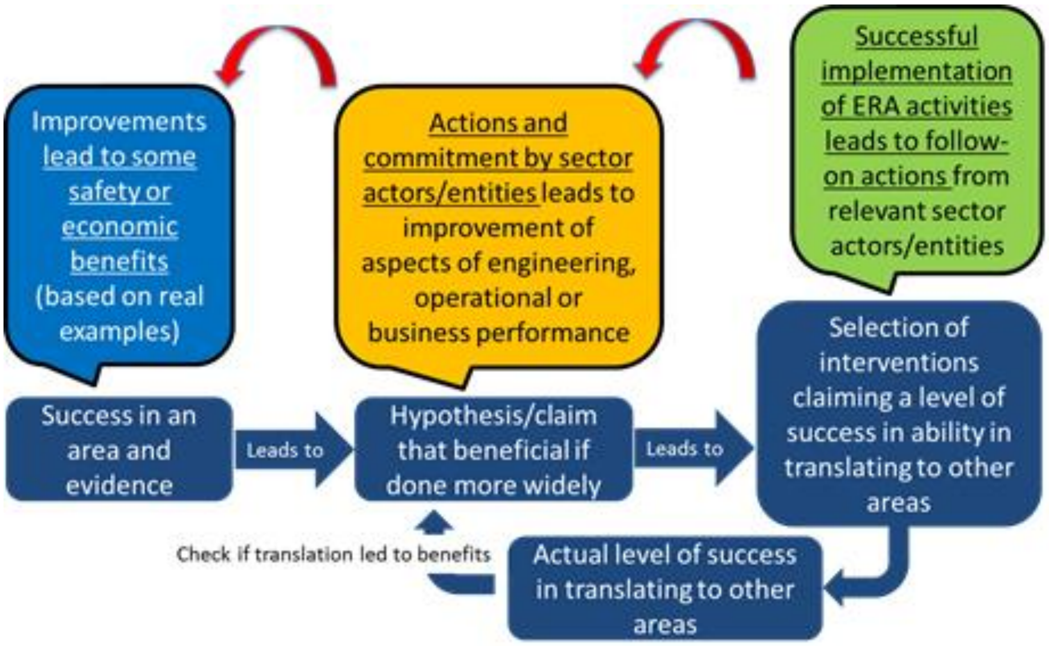

Figure 14: Model of railway system benefits from ERA activities

Once the benefits have been identified in the three stages it is also important to compare them to the cost of the associated changes. These are likely to cover five main areas:

- *Development and implementation of the system and the associated data architecture*
- *Provision and/or collection of accurate and timely data*
- *Changes/redesign of existing systems, data structures and formats including the distributed systems that would input/interact/integrate with the central system*
- *Ongoing maintenance, checking and assurance so that the data being shared and already stored remains fit for purpose*
- *Access to and dissemination of the data outputs with the users.*

In all five areas, costs may be incurred at the:

- *EU and Agency Level*
- *Member state level*
- *Sector area level*
- *Individual organisational level*

It is as essential to understand where the costs are incurred as it is to quantify them as a business case could be desirable at an EU level but may have greater degree of variation in a positive or negative outcome at an organisational level. Some assessment of these aspects is critical to create a comprehensive narrative and fully understand how benefits will be achieved.

Costs estimation

A matrix below allows as a first stage to identify where the material cost implications are likely to be.

Costs categories	**EU level**	**Member state**	**Sector area**	**Organisation/ individual**
1. *Development and implementation of the system and the associated data architecture*	High due to data protections and security requirements	No existing system	N/A	N/A
2. *Provision and/or collection of accurate and timely data*	Moderate as it is assumed that some mechanism/effort to promote and request data will be needed	Moderate	Low	Moderate as the individuals will need to register and provide data
3. Changes/redesign of existing systems, data structures and formats including the distributed systems that would input/interact/integrate with the central system	No impact on existing systems	No impact on existing systems	No impact on existing systems	Minimal
4. Ongoing maintenance, checking and assurance so that the data being shared and already stored remains fit for purpose	Moderate to high due to security and GDPR requirements	Minimal	Low	Minimal
5. Access to and dissemination of the data outputs with the users.	High as user interfaces will need to be sophisticated and user friendly enough to allow for interaction with organisations and individuals	Minimal	Low	Minimal

Note: the scale used none/minimal=*, low/modest=**, Moderate =***, High=****

Aspects highlighted as High or Moderate should be explored further for quantification. However, it is difficult to put a figure on the cost at this stage.

CBA evaluation for improvement cases

Two improvement cases identified by the TF as a priority were evaluated with the use of CBA.

- *Improvement for existing registers for network capacity, and*
- *Improvement of existing registers for last-mile infrastructure.*

In addition, two new use cases, where currently there is not a provision of a centralized register were assessed:

- *A potential register for available and approved drivers if a market of drivers for lease exists, and*
- *A potential register for available vehicles for purchase/lease.*

The quantification methodology used here is aligned with the approach taken by the ESG Task Force on Safety Culture where quantification of the benefits due to the Agency's safety culture projects was provided.

Improvement for existing registers for network capacity

Context

In order to run a train, a railway undertaking first has to ask the responsible infrastructure manager to allocate a path. RU generally either apply for a path well in advance under the annual timetable, or request a path at a later stage on an ad hoc basis among those still available within the reserve capacity. However, the timing set by the infrastructure managers for the construction of the annual timetable is not adapted to the needs of the freight transport sector, as paths have to be booked around 1 year in advance. Unlike passenger traffic, which is more regular and easier to plan, it is difficult for freight operators to anticipate their future demand so far in advance and reserve the most suitable paths available on the network. Rail freight operators, especially smaller ones, are therefore generally forced to use the ad hoc system (for example, this happens in 90 % of cases in Poland). Under this system only a limited number of paths are available (those not yet booked through the annual timetable especially by passenger trains). This often gives rise to one of two situations: either only a limited number of paths are available, with the result that the shipper may have no suitable paths to choose from (so a potential customer uses an alternative mode of transport, usually road transport) or a less suitable path is offered (e.g. a longer, more circuitous route), generally leading to higher costs and more time needed.

Available paths information is nowadays available by IMs, whereas for international path on rail freight corridors (RFC), the IMs have effectively polled this information to RNE (as part of international cooperation). This may cover a major share of paths for cross-border services. The improvement case is about making all path information available, not just those on RFCs in the EU. This would make necessary information available in a more effective and efficient way, which in turns would contribute to a more effective use of available capacity (more paths actually used).

The project TTR (Timetable Redesign) with the expected roll-out for Timetable 2025, should significantly improve the existing situation.

Business use cases

- *Strategic planning (RUs): Identification of prospective routes*
- *Path request (RUs): Availability of paths*

Benefits

Stage 3 - blue box/category (Figure 14) sets out what overall objectives and benefits are being targeted. In the case of database for available vehicles, the objective could be:

• An increase in freight revenues due to better path utilisation:

- *Revenues from freight sales across the EU = 24 060 M€*[5]
- *Given the fact that most of the available path data is made available, this represents a fuller data set being made available as well as a better provision of existing data set. It is reasonable to assume a modest 0.1% increase revenue only due to uncertainties of linking better decision making for freight operators to actual revenue improvements as other market factors will also play a key role. if*

[5] From study: Railway costs and benefits data collection (ERA 2017 38 RS), Final report, Ineco&Ecorys, ERA 2018

freight companies could identify available paths better, a 0.1% increase in revenue is assumed then this equates to = 0.1% of 24 060 M€ = 24 M€ /year

The stage 2 - orange box/ category will identify to what extent a RU's across the EU can use a central database to potentially avail themselves of the benefits highlighted in stage 3. The key considerations are:

• The level of sophistication of the database and its ability to provide real time and robust information to the RU. Rating the different levels of effectiveness for each level of sophistication of a potential database can be as follows (estimates only for illustrative purposes) if a 100% is the ability to fully address the issue:

- *Level 0: The effectiveness is 10% in the absence of any centralised information provision*
- *Level 1: The effectiveness increases to 12%*
- *Level 2: The effectiveness increases to 15% (a 3% improvement over level 1)*
- *Level 3: The effectiveness increases to 18% (an5 % improvement over level 1)*

The current assumption is that for spare network capacity, the existing systems are at level 1 and the desired improvement can take the systems to level 2 or 3 (let us assume in this case the improvement is only to level 3). Therefore, from a potential benefit of 24 M€ if a Level 3 centralised database was provided then it could provide benefits of 1.2 M€/year.

<u>Costs</u>

Costs categories	**Level, costs**
1. *Development and implementation of the system and the associated data architecture*	EU level: 350 000 EUR one-off for IT developments, including common taxonomy development. 300 000 EUR recurring annual costs for further IT development, system support and data maintenance.
2. *Provision and/or collection of accurate and timely data*	Development of interfaces; Data feeding by facility operators for their facilities; Further data collection/update by portal management for facilities/service providers not directly linked to the portal.
3. Access to and dissemination of the data outputs with the users.	Open free access to the web portal, potential fees for additional services (interfaces, reports,…) contributing to a potential future self-sustainable financing.

<u>CBA</u>

	Year 0	*Year 1-20*	*NPV (20 years, 3%)*	*B/C*
Benefits	*0 k€*	*1 200 k€*	*12 650 k€*	*3.65*
Costs	*350 k€*	*300 k€*		

The provision of centrally available rail facilities data is desirable from the economic point of view. This is mainly driven by expected high benefits to the RUs.

Improvement for existing registers for last-mile infrastructure (rail facilities)

Context

All over Europe, rail freight is competing with road transport in numerous market segments. Clients of transport services are demanding transparent and seamless logistics chains on short notice. Therefore, all information needed for the planning and organisation of such customer-oriented rail freight offers must become available instantly and in a user-friendly way. These requirements particularly apply for information on trans-shipment facilities for rail freight transport. In contrast to passenger traffic, such data is – if available at all - distributed amongst numerous public and private owners. The European Commission has identified this lack of a harmonised European data basis and an adequate user tool as a significant barrier for further modal shift to rail freight, especially on cross-border destinations. Against this background DG MOVE procured a study on "User-friendly access to information about last-mile infrastructure for rail freight", which was published in November 2016. The Study included the development of a pilot version of an EU-wide web-based portal with GIS functionalities, capable of presenting relevant data for different kinds of last-mile infrastructure in a transparent way, for a sample of territories in the EU.

In a second study, the existing pilot portal has been further developed towards a "professional" version designed for permanent operation; it is now available under https://railfacilitiesportal.eu/. This version shall become the "tool of choice" that potential rail stakeholders, in particular rail freight stakeholders, should use to access information about rail service facilities in Europe. The Portal provides one of the three ways for service facilities' operators to comply with their obligation under the "Implementing Regulation on access to service facilities and use of rail-related services (2017/2177)" which lays down the details of procedure and criteria to be followed for access to service facilities and services to be supplied in these facilities as listed in points 2, 3 and 4 of Annex II to Directive 2012/34/EU. In the portal the following types of facilities and service providers are distinguished: passenger stations, freight terminal types (Intermodal terminal, multifunctional rail terminal, public siding, private siding), marshalling yard / train formation facility, storage siding, maintenance facility, other technical facility (including cleaning and washing facilities), maritime and inland port facilities linked to rail activities (no separate facility type; location in inland port or maritime port is considered as an add-on to the facility type), relief facility, refuelling facility and service providers without their own infrastructure.

It is notable that the opportunity costs from market failure to deliver an EU-wide (interoperable, or autonomous) solution were not determined. Similarly, benefits of an EU-wide solution were not estimated.

Business use cases

- *Strategic planning (RUs, Transport operators, Logistics service providers): Location, access conditions and availability of last-mile facilities (freight terminals)*
- *Path request (RUs): Availability of facilities and services, compatibility of facilities and last-mile infrastructure with long-haul train path*
- *Train preparation (RUs): Technical conditions of facilities, Location, access conditions and availability of service facilities for maintenance/repair, fuelling, cleaning/washing, etc.*
- *Investors: Planning of new constructions/upgrades/rehabilitation of facilities*

Benefits

Quick access to comprehensive information on all rail service facilities in the EU, presented in a common structure, thereby overcoming the language barrier and assuring higher data quality, reduce costs to RUs. This in particular in international rail freight transport. Besides, the centrally available data allows for a better planning for the physical presence of these services across the EU. The authors of the EC study

"Design features for support programmes for investments in last-mile infrastructure[6]" conclude that "There is a need for investments in last-mile infrastructure of about 9.7 billion € for the period 2015-2030". Some 46% of this total investment need is allotted to intermodal terminals, due to high expected growth rates in this market segment. This means that the foreseen investments for freight terminals alone are estimated at 300 M€ per year.

Stage 3 - blue box/category sets out what overall objectives and benefits are being targeted. In the case of database for available rail facilities, the objective could be to increase the efficiency in service planning, path request and train preparation as well as the increase in the effectiveness of the investments by centrally available data:

• Total average costs for the rail freight sector is estimated as 21 billion €[7]. If it assumed that better provision of last mile infrastructure data to the sector could reduce these costs by 0.25% (over all service provision phases), which equates to 52.5 M€/year (without discounting). This is further reduced by 49 % (domestic rail freight services), to 27 M€/year.

If it further assumed that better provision of last mile infrastructure data to the sector improves the effectiveness of the investments into rail freight facilities by 5%, therefore 300 M€ over a period of 15 years which roughly equates to 15 M€/year (without discounting):

The stage 2 - orange box/ category will identify to what extent a RU's across the EU can use a central database to potentially avail themselves of the benefits highlighted in stage 3. The key considerations are:

• The level of sophistication of the database and its ability to provide real time and robust information to the RU. Rating the different levels of effectiveness for each level of sophistication of a potential database can be as follows (estimates only for illustrative purposes), if a 100% is the ability to fully address the issue:

- *Level 0: The effectiveness is 10% in the absence of any centralised information provision*
- *Level 1: The effectiveness increases to 12%*
- *Level 2: The effectiveness increases to 15% (a 3% improvement over level 1)*
- *Level 3: The effectiveness increases to 17% (an 5% improvement over level 1)*

The current assumption is that for last mile infrastructure, the existing systems are at level 1 and the desired improvement can take the systems to level 2 or 3 (let us assume in this case the improvement is only to level 3). Therefore, from a potential benefit of 42 M€/ if a Level 3 centralised database was provided then it could provide benefits of 2.1 M€/year (1.26 M€/year for level 2)).

Costs

The study "Further development of a European Rail Locations Portal[8]" includes recommendations on the potential future "Business Model" with detailed cost estimations for the development and implementation

---

[6] See https://ec.europa.eu/transport/sites/transport/files/2016-06-rail-final-report-design-features-for-lm-investments-exec-summ.pdf

[7] Study on the Cost and Contribution of the Rail Sector, Steer Davis Gleave, Final Report, 2015 (pg.6) https://ec.europa.eu/transport/sites/transport/files/modes/rail/studies/doc/2015-09-study-on-the-cost-and-contribution-of-the-rail-sector-exec-summary_en.pdf

[8] Report under publication, HaCon 2019, to be available on DG MOVE website later in 2019)

of the portal and permanent operation/management. These are 350,000 € in the first year and 250 000 € in the following years.

Costs categories	**Level, costs**
1. *Development and implementation of the system and the associated data architecture*	EU level: 350 000 € one-off for IT developments, including common taxonomy development. 250 000 € recurring annual costs for further IT development, system support and data maintenance.
2. *Provision and/or collection of accurate and timely data*	Development of interfaces; Data feeding by facility operators for their facilities; Further data collection/update by portal management for facilities/service providers not directly linked to the portal. We assume 24 000 facilities, with data maintenance costs of 2 hours per year (4 hours the first year)) at costs of 27 € per hour[9]. This results in 1.3 M€ for the entire EU (2.6 M€ the first year).
3. Access to and dissemination of the data outputs with the users.	Open free access to the web portal, potential fees for additional services (interfaces, reports,…) contributing to a potential future self-sustainable financing.

CBA

	Year 0	*Year 1-20*	*NPV (20 years, 3%)*	*B/C*
Benefits	*1 680 k€*	*1 680 k€*	*7 100 k€*	*1.3*
Costs	*2 950 k€*	*1 550 k€*		

The provision of centrally available rail facilities data is desirable from the economic point of view. However the efficiency might be negative (B/C<1) if only RU benefits and level 2 data availability is assumed. This is partly due to a relatively high costs associated with data provision for large number of facilities across Europe.

[9] Estimated hourly labour costs, 2018, Eurostat, https://ec.europa.eu/eurostat/statistics-explained/index.php/Hourly_labour_costs

A potential database for available and approved drivers

Context

When an RU consider expanding its services, faces shortage of train drivers, needs to assure service contingency, they may be a need of engaging a driver licensed and certified for a given route. There might be train drivers engaged with other RUs, but still available for parallel, or punctual engagement.

Business use cases

- *Strategic planning (RUs): Identification of potentially available drivers*
- *Path request RU: Identification and engagement of an available driver*
- *Train operation: In case of contingencies, identification of an available driver*

Benefits

Stage 3 - blue box/category sets out what overall objectives and benefits are being targeted. In the case of database for licenced and available train drivers, the objective could be:

- A reduction in the cost of securing the services of a driver where the key data set could be:
  - *Total number of full time equivalent (FTE) train drivers across the EU = 100 000 drivers (RMMS[10])*
  - *Train driver licences issued by NSAs = 153 000 (ERA[11])*
  - *Potential number of train drivers available = 153 000 -100 000 = 53 000*
  - *Potential number of train drivers who could be recruited/engaged = Worst case scenario - 1% of 53 000: 530 drivers*
  - *Cost of training saved by employing an already licenced and available driver instead of recruiting someone new to train driving and without a licence = 10 000 € (this is a dummy estimate used for illustrative purposes and requires further validation).*
  - *Potential for cost reduction if an RU could identify and recruit a driver quickly to satisfy need: 5.3 M€*

The stage 2 - orange box/category will identify to what extent a RU's across the EU can use a central database to potentially avail themselves of the benefits highlighted in stage 3. The key considerations are:

- The level of sophistication of the database and its ability to provide real time and robust information to the RU which it can act upon to recruit an available driver. Rating the different levels of effectiveness for each level of sophistication of a potential database can be as follows (estimates only for illustrative purposes) if a 100% is the ability to fully address the issue:
  - *Level 0: The effectiveness is 10% in the absence of any centralised information provision*
  - *Level 1: The effectiveness increases to 12%*
  - *Level 2: The effectiveness increases to 15%*
  - *Level 3: The effectiveness increases to 20% (a 10% improvement over no database)*

---

[10] Fifth report on monitoring developments of the rail market, EC, 2018

[11] Report on Railway Safety and Interoperability in the EU (2018), ERA, 2018

Therefore, from a potential benefit of 5.3 M€ if a Level 3 centralised database was provided then it could provide benefits of 0.53 M€ per year.

The stage 1 - green box/category sets out short-term outcomes i.e. the effectiveness of the database in delivering against its specification and achieving level 3 functionality:

Let us assume that there is an 80% chance of delivering a database which could achieve Level 3 functionality and work effectively.

Therefore the benefits a database could provide would be in the order of 425 000 €.

Costs

Costs categories	**Level, costs**
1. *Development and implementation of the system and the associated data architecture*	EU level: 200 000 EUR one-off for IT developments, including common taxonomy development. 50 000 EUR recurring annual costs for system and data maintenance. The protection of private information to account for majority of the costs.
2. *Provision and/or collection of accurate and timely data*	Recurring data input required from single drivers: 53 000 drivers, one entry per year at 9 EUR per entry > 500 000 EUR
3. Access to and dissemination of the data outputs with the users.	Possibly paid access to make the system financially autonomous.

The critical elements is the usage of the system by the drivers. Using the assumptions from the previous section on benefits, and assuming that 50% of potentially available drivers introduce their data.

CBA

	Year 0	*Year 1-20*	*NPV (20 years, 3%)*	*B/C*
Benefits	*0 k€*	*425 k€*	*-2 500 k€*	*0.70*
Costs	*700 k€*	*550 k€*		

The potential centralized database for available approved drivers is not economically viable, mainly due to the high data maintenance costs.

On the other hand, a centrally available data on approved drivers might be cost effective, since the data are already available (by single NSAs). This despite a rather modest expected benefits from improved safety assurance.

A potential database for available vehicles

<u>*Context*</u>

When an RU consider expanding its services, a timely availability of rolling stock is crucial. This in particular in passenger train service operations. Relying on a used rolling stock owned but other parties, but not used for operation, is economically sound and quicker alternative to the acquisition of new rolling stock. The availability of coaches/train sets to third party is usually known in advance (2-3 years) and fit well in the strategic planning timetable of RUs considering the expansion of their services.

<u>*Business use cases*</u>

- *Strategic planning (RUs): Identification of potentially available vehicles*

<u>*Benefits*</u>

Stage 3 - blue box/category sets out what overall objectives and benefits are being targeted. In the case of database for available vehicles, the objective could be:

• An increase in the probability of securing revenue from renting rolling stock where the key data set could be:

- *Revenues from rolling stock rentals (RUs and Leasing companies) = 5 300 M€[12].*
- *Potential 0.1% increase in the probability of realising that revenue if RUs could identify available vehicles which are compatible with the relevant area of use or routes or markets = 0.1% of 5 300 M€= 5.3 M€*

• Plus additional benefits by reducing the cost of identifying a compatible and available EMU or DMU where the key data set could be:

- *Total number of EMUs and DMUs = 78 000; total number of coaches = 320 000*
- *Potential number of EMUs or DMUs available for re-circulation which may be compatible with area of use an RU is considering= 0.5% of 398 000 = 1 990*
- *Cost of identifying an EMU or DMU which is compatible with relevant area of use = 5 000 € per vehicle (this is a dummy estimate used for illustrative purposes and requires further validation).*
- *Potential for cost reduction if an RU could identify an available vehicle quickly to satisfy need= 1 990 x 5 000 € = 10 M€*
- ***Total potential benefits of 5.3 + 10 M€ = 15.3 M€***

The stage 2 - orange box/ category will identify to what extent a RU's across the EU can use a central database to potentially avail themselves of the benefits highlighted in stage 3. The key considerations are:

• The level of sophistication of the database and its ability to provide valid and robust information to the RU which it can act upon to identify an available EMU or DMU. Rating the different levels of effectiveness for each level of sophistication of a potential database can be as follows (estimates only for illustrative purposes) if a 100% is the ability to fully address the issue:

- *Level 0: The effectiveness is 10% in the absence of any centralised information provision*

[12] Railway costs and benefits data collection (ERA 2017 38 RS), Final report, INECO&ECORYS, 2018 (Pg.27)

- *Level 1: The effectiveness increases to 12%*
- *Level 2: The effectiveness increases to 15%*
- *Level 3: The effectiveness increases to 20% (a 10% improvement over no database)*

Therefore, from a potential benefit of 15.3 M€ if a Level 3 centralised database was provided, then it could provide benefits of 1.53 M€ per year.

The stage 1 - green box/category sets out short-term outcomes i.e. the effectiveness of the database in delivering against its specification and achieving level 3 functionality:

Let us assume that there is an 80% chance of delivering a database which could achieve Level 3 functionality and work effectives.

Therefore the benefits a database could provide would be in the order of 1.22 M€ per year.
Costs

Costs categories	Level, costs
1. *Development and implementation of the system and the associated data architecture*	EU level: 300 000 EUR one-off for IT developments, including common taxonomy development. 100 000 EUR recurring annual costs for system and data maintenance. The protection of private information taken into account.
2. *Provision and/or collection of accurate and timely data*	Recurring data input required from single vehicle owners: 1 990 units per year at 27.5 EUR per data entry costs (one hour), updated annually at the same costs. Total yearly costs 55 000 EUR.
3. Access to and dissemination of the data outputs with the users.	Possibly paid access to make the system financially autonomous.

It is assumed that 50% of all EMU/DMU potentially available for lease (780 assumed in benefits section) would be introduced in the database.

CBA

	Year 0	*Year 1-20*	*NPV (20 years, 3%)*	*B/C*
Benefits	*0 k€*	*1 224 k€*	*15 100 k€*	*6.67*
Costs	*355 k€*	*155 k€*		

The provision of centrally available rail facilities data is desirable from the economic point of view. This is mainly driven by expected high benefits to the RUs and relatively low recurring costs.

6.4. Consolidated assessment results

The results of the qualitative and quantitative assessment for improvement cases are compared, showing a good alignment for three of four analysed cases. In case of available second-hand vehicles, the quantitative assessment yields highly positive results, compared to the qualitative a slightly negative assessment carried out by the expert group.

Improvement case	B/C - expert group	B/C - CBA
Network/route capacity	1.53	3.65
Rail facilities	1.24	1.30
Available approved drivers	0.85	0.70
Available compatible vehicles	0.90	6.67

These results also suggest that expert qualitative assessment could yield valid results, in particular when the full quantitative assessments are not deemed to be proportionate or if data is missing or unavailable. The combination of a rough quantitative assessment to supplement expert judgement can further validate the judgement, thereby narrowing the cases where further full quantitative assessment is necessary and justifiable.

7. Findings

7.1. Data use cases

a. Business use-case description for data in train operation is largely absent in all EU strategic and legal documents and leads to shortcomings in the design of IT systems and in railway system data management.

b. The fitness check of existing Agency-maintained registers/databases and the RU user needs identify a number of opportunities for enhanced data integration. The majority of them are however in the area of operational data, where Agency can bring only a limited value.

c. For the data types with potential for EU wide data sharing, only approved drivers and safety occurrences belong to administrative/regulatory data. Any action leading to a development of a new database would have to be justified by an impact assessment (being out of scope of this report). For the remaining absent data types, a support may need to be provided by the Agency, to the sector, in the form of common standards development (reference data).

d. Economic assessment of business use cases relying on IT systems development is a challenging and resource consuming tasks made possible by an adopted methodological framework for CBA assessment. The improvement cases assessed in this report show that there is an economic case for systematic centralised sharing of infrastructure-related data (available paths and rail facilities) and for sharing of data on available second-hand vehicles.

e. Although subject to qualitative economic assessment only, the use cases linked to the use of infrastructure data, notably the route book, a description of infrastructure compliant with ATO, the network capacity and the transient infrastructure were found to be of the highest value to the RUs. Among them, the implementation of an effective system for vehicle/train – area of use/route compatibility assurance is of the highest value to the RUs.

7.2. Data management

a. A high-level review of taxonomies in existing registers/databases maintained by the Agency points to a poor railway master data management (MDM) and of railway reference data management (RDM) at the EU level. (It was notably emerging from RU). In order to be efficient, it needs to be driven by an authority such as ERA, instead of being let to the sector and private entities.

b. Some data elements available in databases and registers are not publicly available, although they are publicly available at the national level (e.g. Swiss VVR data). Some other data elements are only available through a specific user interface, with a valid registration. Making more SERA data open have a potential to provide for new opportunities, knowledge and innovation in rail. Experience from several MSs suggest that Open data concept applications lead to more efficient and effective railways.

c. Throughout the interviews with RUs, carried out by TF members, it became obvious that they are not as much concerned with the availability of the data in existing registers and databases, but their functional availability (usability). The approach to data quality, in particular within regulatory databases and registers, was felt as insufficient and the degree of data interoperability as too low.

d. Whereas the data quality aspects were identified as crucial in the context of EU railway data management, with the following data quality dimensions were put forward: consistency, coherence and timeliness.

e. Various data interfaces have been developed for rail data and used in various ways in different databases. Notably, the TAF TSI CUI is a railway specific solution developed mainly by the incumbent operators. To facilitate an efficient and standardised data exchange within the railway sector for all stakeholders, a further development of the TAF TSI Common Interface (specifications) compatible with industry-wide universal CUIs, such as Soap web service, may be required.

7.3. Legal framework

a. Content of the vehicle marking is prescribed by TSI OPE. An update is necessitated following an upgrade, or update in ownership. Due to operational reasons (repainting will only be done during the next workshop visit), the marking may become temporarily outdated. Could it perhaps be replaced by an EVN?

8. Recommendations

Below are the recommendations to the Agency and to the Commission. They are grouped topically and not sorted by importance.

8.1. Vision/target system

a. Define a target system, from which a data architecture developed in this report could be refined. It should include a roadmap for rail open data and common standards (incl. CUI) and taxonomies development.

b. Make use of a CBA to determine the value of data in registers/databases maintained by the Agency and justify their further developments motivated by use cases determined in this report.

c. Economic assessment CBA for all (relevant) improvement cases: Qualify value of data in single db/registers, where possible (e.g. stated preference from RUs on improvement cases/ new developments).

d. Systematically involve current and future users served by registers/database at an early stage of their development.

e. Take into consideration the latest IT trends in planning of future IT systems/databases as to assure efficiency and technical compatibility as well as interoperability with other systems.

8.2. Improvements in existing databases

a. Registers/databases hosting infrastructure data are top priority for further development as to satisfy, in a most possible efficient way, the business needs of the RUs. These are, in the order of priority: Route-vehicle compatibility, Path planning, Train running, Route book and the ATO/CCS. It is notable here, that existing VVR and RINF were developed under the assumption of high benefits from route-vehicle compatibility assurance they enable.

b. Particularly for the infrastructure description in RINF, it is proposed to focus as target infrastructure description on the granularity needed for ATO, this granularity being compliant with the infrastructure related use cases.

c. Make an inventory of purposes of single registers/db and check against the common use cases (business goals) determined in this report. Systematically monitor the extent of their fulfilment. Derive data model (on data elements) as a basis for a systematic MDM.

d. Systematically rationalize the reporting and consumption of market data, by designating master data for each area and make them easily available. Prioritize the development of data architecture for safety occurrence data and ERTMS data, built on the identified regulatory and operational use cases.

8.3. New databases developments

a. Creating a centralized database of available rolling stock appear to be desirable from the economic point of view. It may facilitate new rail services development and strengthen the competition in the rail passenger transport.

b. Sharing of regulatory data on train drivers with EU license should be privileged over a more operational data on train driver availability, although this could change in close future due to growing concern on train driver concerns in some countries.

8.4. EU railway data management

a. Strengthen existing Master and Reference Data Management of rail regulatory data, with the Agency leading the effort. The Agency should notably take up the role of a rail data authority, pro-actively working with other parties.

b. The Agency should develop an inventory of all data categories/objects used in all rail administrative/regulatory registers/databases, to identify inconsistencies, identify candidates for master and reference data management and for improvements in the legal framework.

c. Data quality assurance mechanisms should be actively implemented in existing Registers/databases managed by the Agency. For example, automated data quality checks are needed to compare inconsistencies in data in various databases.

d. Assure timeliness of the data by introducing predictability in system/taxonomies updates (stepped process, baselining) and by data collection assurance mechanisms. This is closely linked to data quality assurance for data in registers and databases hosted by the Agency.

e. Roles and responsibilities are now defined at the level of single databases/registers, in practice in their underlying legislation, while a more holistic approach is advocated, stemming from the generic roles and responsibilities of actors under EU rail legislation.

8.5. Legal framework

a. Assure full alignment in ERATV/ TSI-OPE, RINF/TSI-OPE data attributes, and also CCS as far as the infrastructure description for ATO is concerned.

b. Review roles, responsibilities and timeliness for data provision as to meet purposes of single tools. For example, the vehicle registers does not facilitate use of vehicle data for route-vehicle compatibility check, or use of data for statistical purposes.

c. Introduce enforcement mechanisms to support quality and timeliness of data in regulatory registers/ databases.

9. References

BABOK©: A guide to the business analysis body of knowledge V.3, International institute of business analysis, 2015

DAMA- DMBOK©: Data management body of knowledge V.2, DAMA international, 2017

Annex A: Task force members and observers

MEMBERS:

Name	Organisation
Vaibhav PURI	NSA UK / RSSB
Paweł RAFALSKI	NSA PL
Thomas HEYDENREICH	UIP
David GILL	Network Rail
Christian MÜLLEN	DB Netz
Javier MORENO	EIM
Jacques AUDENAERT	EIM
Markus RIEDER	OBB / CER
Peter TAUSCHITZ	OBB / CER
Richard JOHN	OBB / CER
Vojkan STEFANOIC	RNE
Niklas GALONSKE	HACON
Christian WEBER	SNCF / CER

OBSERVERS:

Name	Organisation
Andreas Von Wuellerstorff	UNIFE
Yann SEIMANDI	CER