ERTMS Challenges for a Safe and Interoperable European Railway System

Katja Schuitemaker
Design, Production and Management Department,
University of Twente
Enschede, the Netherlands
email: k.schuitemaker@utwente.nl

Mohammad Rajabalinejad
Design, Production and Management Department,
University of Twente
Enschede, the Netherlands
email: m.rajabalinejad@utwente.nl

Abstract—The European Railway Traffic Management system (ERTMS) aims at replacing the different national train control and command systems in Europe and will serve to make rail transport safer. In order to provide insight into safety developments within the European railway system, the present study evaluates ERTMS at both national and international integral level. For this purpose, the international data from European ERTMS implementations is combined with national data obtained from interviews with Dutch ERTMS stakeholders and safety experts. Effects of deregulation, dynamic specifications, interoperability and time drain make that allowing an interoperable railway system by implementing ERTMS appears not to be self-explanatory. Also, without an overarching process, cross-discipline understanding and improved ascribing meaning to data, implementing ERTMS does not mean the railway system will become safer.

Keywords - ERTMS; safety; integral assessment; socio-technical safety.

I. INTRODUCTION

As early as the 1990s, the European Commission (EC) decided passenger trains should be able to travel across international borders in Europe. In 1998, the EC requested the foundation of the Union Industry of Signalling (UNISIG) and assigned this with the task of drafting technical specifications of the European Railway Traffic Management System (ERTMS). ERTMS was designed to be fully interoperable across the European Union and has become the European standard for the Automatic Train Protection (ATP) that allows one to drive European trains across the European Union and will serve to make rail transport safer [2]. Some explanations on why ERTMS is considered to increase railway safety include:

- Continuous supervision of the train speed [2]. This means that the train can receive authorization to continue running at maximum allowed speeds continuously through the GSM-R system (only available at ERTMS level 2).
- Reduce the risk for human errors [3]. For example, work related errors caused by stress, sleepiness, fatigue, and sleep disturbance.
- Decreasing the amount of Signals Passed At Danger (SPAD) [4]. This can be explained by the difference with previous ATP systems. ERTMS is able to influence trains driving under 40 km/h.

However, it appears that implementation of ERTMS does not automatically mean a safer railway system. For instance, in practice, in the Netherlands, the amount of SPADs under ERTMS (both Level 1 and Level 2) was 11 in 2013, and 17 in 2015 [5]. Though, these numbers are low, so they can be considered as an indication, not necessarily as a trend. More studies, both scientific and industrial, question the safety level through implementation of ERTMS.

At the international level:
- Smith et al. addresses issues relevant to safe introduction of ERTMS into European railway systems [6]. These issues include technical system integration, technical system failures and human factor considerations.
- Laroche and Guihéry study the European Transport Policy, the role played by the EC, the ERTMS innovation process in accordance with innovation process in surface transport, and the difficulties for the implementation of an intelligent transportation system innovation [7].
- Ghazel addresses the regular evolving documents [8].
- The EC itself has studied past and current problems with ERTMS implementation [9].

At the national Dutch level:
- The Ministry of Infrastructure and the Environment, ProRail and NS have collected (im)possibilities of ERTMS [10].
- ProRail and NS executed a pilot for gaining experience with driving under ERTMS [11].
- A specialized team investigated the sequence of events and decision making processes in the Netherlands, which have led to delays in deployment of the ERTMS train signalling systems in the HSL railway project [12].

Most importantly, ERTMS principles have led to assumptions regarding safety. This paper studies safety implications of ERTMS at the integral level, questioning:

Assumption 1: the fact that ERTMS aims at replacing the different national train control and command systems in Europe [13]. The risk comes from the idea that ERTMS is considered to be a replacing system, instead of a new system with new interfaces in itself.
• Assumption 2: the fact that ERTMS will serve to make rail transport safer [2]. This is not self-evident. Parties may gradually sail closer to the wind, thereby unintentionally and unnoticed, compromising too much on safety. Only when things go wrong - as in the Hilversum derailment - it becomes clear that a threshold has been passed [14].

Section 2 provides an overview of the background of railway deregulation, ERTMS specifications, European interoperability, and cost reduction as a result of a change in organisational behaviour. The methodology is discussed in Section 3. Section 4 explains findings with regard to the safety architecture and sociotechnical safety of ERTMS. Findings are discussed in Section 5. Section 6 summarizes, concludes and highlights challenges.

II. BACKGROUND

The railway system includes technical, managerial, organizational, and regulatory aspects. The subsystems can work perfectly individually, but together they can create a hazardous state. Many factors, both technical and socio-institutional in nature should be combined to turn a serious challenge of one European train system into a great success satisfying social needs of lower costs, better utilization of an infrastructure and less complex logistics [15].

A. Deregulation

Starting with the 90s, in order to promote greater competition, the rail industry in Europe was restructured. On one hand, the vertical separation means management and ownership of infrastructure are totally separated from other rail activities. On the other hand, various operators are using infrastructures. Deregulation is the reduction or elimination of government power in industry, usually enacted to create more competition within the industry. In addition, safety regulation has increased a hundred-fold between 1947 and 2008 [16]. At last, the ERA explains a shift from quantitative safety data to qualitative safety data [17].

The privatization and deregulation has led to an increased involvement of private actors, national and international [18].

B. ERTMS specifications

The Union Industry of Signalling (UNISIG) was founded in 1998/99 at the specific request of the European Commission (EC) [19]. It was created to develop ERTMS specifications. The final version of ERTMS specifications is published by the EC following the approval of the Member States.

In November 2012, the EC intentionally deleted ERTMS Functional Requirement Specifications making these specifications no longer mandatory. The remaining System Requirements Specifications are written in natural language, which allows multiple interpretations [20].

C. Interoperability

The meaning of interoperability is two-fold. On one hand, interoperability refers to a geographical interoperability between countries and between projects. On the other hand, it also refers to interoperability between suppliers. This opens the supply market and increases competition within the industry [21]. The result of this is the absence of a single entity that is responsible for the railway system as a whole.

D. Cutting cost and time

Dutch national safety goals are approached through use of the As Low As Reasonably Practicable-principle (ALARP) and standstill-principle [22]. For risks in the “ALARP area”, all potential risk reducing measures must be evaluated in terms of cost efficiency, cost-benefit balance or some similar economic measure. Finally, selected risk-reducing measures may be introduced based on experience or best practice in combination with cost-efficiency considerations [23].

According to Rasmussen [24], systems and organizations continually experience change as adaptions are made in response to local pressures and short-term productivity and cost goals. Several accidents such as Bhopal, Flixborough, Zeiburg, and Chernobyl demonstrate that they have not been caused by a coincidence of independent failures and human errors, but by a systematic migration of organisational behaviour toward accident under the influence of pressure toward cost-effectiveness in an aggressive, competitive environment. There happens to be a standing request to be cost-effective in risk management [25]. According to the ERTMS strategy group of Great Britain, initially the principal requirement for ERTMS was to improve safety. “Over approximately the last ten years, capacity became a more significant influence and then, more recently, cost reduction [26].”

III. METHOD

In order to investigate the nature of phenomena, a qualitative approach in the form of interviews was executed. The findings in this paper are based on international data from European ERTMS implementations linked with national data obtained from semi-structured interviews with Dutch ERTMS key stakeholders and safety experts from train operating companies, infrastructure managers and self-employers involved with the ERTMS national program. In total, 15 semi-structured interviews have been held, performed face to face, lasting between 30 and 90 min. All interviews were audio recorded and transcribed verbatim.

Emphasis was placed on the ERTMS safety architecture and on social technical safety of ERTMS on both Dutch national and international level. The topics discussed included effects from deregulation, ERTMS specifications, interdisciplinarity and time drain.

Transcriptions were processed through qualitative inductive content analysis in order to develop a theory and identify themes through repeated examination, comparison, abstraction and data reduction. The material was abstracted and reduced to a set of themes. The procedure was repeated to refine chosen themes. Two main categories were identified as a thread through transcriptions: (1) implications with regard
to the safety architecture, (2) implications with regard to socio-technical safety.

IV. FINDINGS

A. Safety architecture

With the implementation of a single signalling system through Europe, the EC has opted for radical innovation for all Member States. Similarly, engineers are often in favour of the most innovative, not yet proven technology [27]. At the same time, instable specifications make it difficult to adopt a radical innovation [7] and issues occur with adapting the new system to the old one. Once a hazard scenario is identified, it is not trivial to identify all the possible causes in the system [28]. In other words, a system that is new, or particularly complex can generate scenarios that are not included in the standard set.

Earlier studies explain that ERTMS specifications are instable [7], [26], written in informal language [8], non-consolidated [6] and incomplete [15], [27]. Up until today, stakeholders indicate specifications are not sufficient. To be more specific, missing parts concern management, integral system integration and physical design.

As a result, updates are postponed in anticipation of new specifications, covering multiple requirements through one update.

For specifications, preferences vary on both international and national level. The signalling system for the trajectory the Netherlands – Germany (remote monitoring) differs significantly from the signalling system for the trajectory the Netherlands – Belgium – France (more autonomy for the train driver, missing track signalling), which is more in line with ERTMS Level 2. Therefore, to migrate to ERTMS Level 2, France does not have to change much. To migrate to ERTMS Level 2, Germany and the Netherlands face a discontinuation with the past. The signal systems as well as the automatic train protection systems are still different from one EU country to the next [28]. In addition, the various ERTMS levels include different technical requirements and applications. A higher level involves less track side equipment, but more on-board equipment. This change also implies that the costs of the signalling system will migrate from infrastructure managers to train operators [12]. Infrastructure managers might anticipate on the developments, where operators do not like to upgrade existing rolling stock [29].

In the same line, various subsystems of the railway system are tendered. At the national level, the 5 ERTMS-projects are explained in table 1.

<table>
<thead>
<tr>
<th>Project</th>
<th>Supplier</th>
<th>ERTMS level</th>
<th>In service date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Betuweroute</td>
<td>Alstom</td>
<td>2</td>
<td>2007</td>
</tr>
<tr>
<td>Port Rotterdam</td>
<td>Alstom</td>
<td>1</td>
<td>2009</td>
</tr>
<tr>
<td>High-Speed Line South</td>
<td>Thales/ Siemens</td>
<td>1/2</td>
<td>2009</td>
</tr>
<tr>
<td>Lelystad-Zwolle railway</td>
<td>Alstom</td>
<td>2</td>
<td>2012</td>
</tr>
</tbody>
</table>

Various ERTMS levels result in multiple transitions. Table 2. shows all the 21 possible transitions between ERTMS levels [31].

<table>
<thead>
<tr>
<th>From</th>
<th>To</th>
<th>STM</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>STM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In practice, the choice for no nationwide rollout of one ERTMS variation results in many transitions between various subsystems. In other words, realization of implementation is unique for every project, and dependent on stakeholders, environment and activities. As is also explained by Leveson [31], the interconnectivity and interactivity between system components make that greater complexity leads to vastly more possible interactions than could be planned, understood, anticipated or guarded against. In reality, Table II shows just a fraction of the number of transitions. As is also concluded by Smith [6], existence of many ETCS versions with technical problems require the need for a backup system.

As a result, systems can be incompatible, for example, the two implementations made by Alcatel (Dutch part of the railway) and Alstom (Belgium part of the railway).

In the end, with ERTMS, complexity of technology, use, and processes of the railway system increases. Interviewees indicated that the technological development in ERTMS is underestimated. A failure with ETCS can have up to 100 causes where train drivers and signalmen must find a solution through difficult procedures and processes, with limited technical system knowledge.

B. Sociotechnical safety

Deregulation has led to considerably more actors on the market. As is also concluded by [6], incomplete/unstable specifications of ERTMS are further hampered by companies involved. With Dutch automatic train protection (ATB) tracks, only one manufacturer (Alstom) was involved. With ERTMS and the tendering of subsystems, various manufacturers are involved. Stakeholders within the ERTMS program come from, among others operating companies, infrastructure provider and self-employers.

In the first place, the rising number of parties involved entails a considerable amount of points of view, skills, responsibilities and interests to the interaction. Boundaries of what constitutes the system become fuzzy; interdependencies and interactions multiply and mushroom [32]. With so many interests, there is a risk of compromising too much on safety. In this context, the train derailment in Hilversum teaches us that the related interests can gradually and unnoticed apply pressure on the management of safety risks [14].
Second, qualitative data uses subjective indexes and is based on logical reasoning from multiple experts. Differences in both language and culture can be major barriers to multidisciplinary work [33]. On top of this, countries are using their own language, making intersectional issues even more complex. For these reasons, on both international level and national level, stakeholders experience difficulties with understanding their respective system. This is also described by Forsberg [3], who states that the new societal organization indicates that intersectional issues and decisions have increased between the various actors, particularly since mishaps or accidents often are caused by circumstances or weak links between them.

Then, for interoperability, without legal entity, it is difficult for the national ERTMS program to allocate responsibility of interface risks to a stakeholder. What makes this even more difficult, is a missing central designer, or any party, that knows the entire complex system. As is also described by Baxter [33], borders between disciplines have been largely maintained despite efforts at creating interdisciplinary teams by involving domain specialist in the design process. One discipline does not fully understand what other disciplines can do, because it essentially stops after collecting data rather than analysing data to ascribe meaning to it so that it could be more readily used by others. Although the European exposed Common Safety Method (CSM) aims at an integral safety approach, the final report on the ERTMS pilot between Amsterdam and Utrecht explains that “overarching processes between railway and train transportation are missing and that these are necessary for optimum implementation of ERTMS”. Employees are often focused on their own job, knowing a lot about their own subject. In practice, data is set and sent to the next. Organizations feel responsible for their own processes, not for the integral railway system as is also recognized by the Dutch Ministry of Infrastructure and the Environment [11].

As it is also described by Nusser [34], “Black box” approaches are regarded with suspicion – even if they show a very high accuracy on the available data – because it is not feasible to prove that they will show a good performance on all possible input combinations. This makes it possible to make a decision based on hidden factors. For example, as safety is sometimes seen as hindrance to effective marketing solutions, focus can be on finishing on time and approve of design.

This is also questioned by Enserink: “It is strange to see how in many large projects, such as the Westerschelde tunnel, the Betuwelijn, and the ‘Groene Hart’ bored tunnel of the high speed rail line south, the discussions of safety issues and safety management took place at a very late stage in the project cycle” [35]. He also explains: In all the examples in the planning phase, the analysts neglected the safety issues or these issues were temporarily stalled because of their complexity. Many have been investigated with the aim of asportioning blame or liability and although safety recommendations are often made, they frequently fail to identify some of the underlying causes of whatever went wrong [28].

As for time drain, with such a large number of stakeholders, it is not always possible to involve every stakeholder in substantive discussions. As a result, stakeholders try to represent another’s perspective. In addition to this, stakeholders are involved in multiple projects resulting in conflicting goals. In other words, the decision-making process can be person related, instead of organization related. According to the parliamentary commission Fyra, it seems like safety has become a subject for negotiation [36]. In practice, the safety case HSL-Z resulted in only mitigating major issues due to time pressure [37].

In the end, any of these effects enable local actors to change their conditions in one of its corners for a very good reason without apparent implications. This can bring immediate gains on some local goal trade-off. Both Leveson [31] and Dekker [32] explain that with a vast number of widely distributed interacting components in an organization, small ‘drifts’ in procedure or policy will not necessarily be identified as risks to the safety of the SoS. Figure 1 shows the interrelationships between the effects of deregulation, dynamic specifications, interoperability and time drain involved with ERTMS.

V. DISCUSSION

A. Safety architecture

Both internationally and nationally, stakeholders have different preferences for ERTMS design. In addition, specification interpretations by manufacturers vary. This in turn leads to a system of a wide variety of subsystems and the associated increase of transitions. This, whilst the ERTMS program tries to prevent transitions between subsystems [38]. The checking of the critical specifications in the natural language is a burdensome task. At the same time, new trains are designed, without required ERTMS specifications. In practice, the instable specifications and various interpretations are a major issue when dealing with such systems. The consequences are significant: the 5 ERTMS-projects (Betuweroute, Port of Rotterdam, HSL South, Amsterdam-Utrecht and Lelystad-Zwolle) are all different [10], let alone the European variations.

Both international and national preferences, changing specifications, changing stakeholders and varying manufacturers led to a unique realization for every subsystem. Along the same line, the occurrence of further transitions with accompanied complexity and procedures and processes that multiply and have wider ramifications.

To pick up from the assumption that ERTMS replaces the national train control and command systems, all in all, it can be concluded that the greater complexity leading to vastly more possible interactions, was and is, unforeseen. As of today, the goal for one interoperable railway system is not achieved. Challenges concern systems thinking accompanied by complex interdisciplinary systems.

B. Sociotechnical safety

With ERTMS and the tender strategy, information has become more sensitive. As a result, stakeholders lack insight into cross-border information. Also, ERTMS involves an
increased number of stakeholders that differ in both language and culture. As for safety, this means that lack of availability of information makes it difficult to determine a root cause.

A central designer that knows or has the responsibility over the entire complex system misses. This, in combination with stakeholder involvement in multiple projects, make local-goal trade-offs possible. Since there is no integral view, local actors can change their conditions without apparent implications.

Time drain and pressure towards cost-effectiveness can inadvertently lead to generating adaptive responses, wrong/missing identification of hazards and safety risks and also to safety concessions. In the same line, if only catastrophic issues are addressed, many other hazards may go uncorrected, which may have a costly impact.

To pick up from the assumption that ERTMS will serve to make rail transport safer, both implicit data-exchange and a missing integral view make it hard to perform a comprehensive safety assessment. Challenges concern the overarching process, cross-discipline understanding and ascribing meaning to data.

VI. CONCLUSION

In present study, the effects of deregulation, dynamic specifications, interoperability, and time drain on the European railway system have been researched.

First, allowing an interoperable railway system by implementing ERTMS appears not to be self-explanatory. Second, implementing ERTMS does not mean the railway system will become safer.

Specifications allowing multiple interpretations result in various design choices, disparities between systems, possible little recognition of hazards and risks, and cumbersome procedures. Realizations are dependent on stakeholders, environment and activities. As of today, the goal for one interoperable railway system is not achieved. Challenges concern systems thinking accompanied by complex interdisciplinary systems. A missing central designer and overall process lower the degree to which the parties succeed in correctly harmonizing various processes. Since there is no integral view, local actors can change their conditions without apparent implications.

As is also concluded by the Dutch Safety Board [39], railway undertakings should make transparent as to why they decide to implement certain measures or not. Challenges
concern the overarching process, cross-discipline understanding and ascribing meaning to data.

REFERENCES


