Conférence Européenne des Directeurs des Routes

Conference of European Directors of Roads

PRESORT

ImPRoving thE uSe Of third-paRTy data by NRAs (PRESORT)

Report on Deep Dive

Deliverable D4.1 Version 1.1

Date 11 March 2025

CEDR Call 2022 Data: Maintaining and sharing the digital road infrastructure

ImPRoving thE uSe Of third-paRTy data by NRAs (PRESORT)

D4.1 Report on Deep Dive

Due date of deliverable: 31 July 2024

Actual submission date: 30 October 2024

Resubmission date: 11 March 2025

Start date of project: 1st May 2023

End date of project: 31st May 2025

Author(s) of this deliverable:

Giovanni Huisken, MAPtm

Xiaoyun Zhang, MAPtm

Version: 1.1

Executive summary

This report, D4.1 Deep Dive of Work Package 4 (WP4), provides insights into the practical use of third-party data by National Road Authorities (NRAs) across Europe. By combining workshops and desktop research, it investigates practices and experiences to generate use case findings and outcomes. It concludes with lessons learned regarding the procurement, implementation, and use of third-party data by NRAs. The outcomes of this report offer NRAs guidance and advice for utilising third-party data and solutions at various stages, contributing to the overall advancement of the PRESORT project's goals.

Previous results from WP2 and WP3 highlight the significant opportunities third-party data present for NRAs, but also underscore the challenges they must navigate. These include procuring cost-effective data, ensuring data quality, integrating third-party data with in-house systems, monitoring Service Level Agreements (SLAs), and adhering to regulatory standards, among others.

The primary objective of D4.1 is to provide an in-depth overview of the outcomes and lessons learned by the selected NRAs in procuring, implementing, and utilising third-party data. The sub-objectives aim to address the 'what,' 'how,' and 'so what' of the primary objective.

To meet these objectives, D4.1 employed four key building blocks:

- Desktop research, to identify knowledge and gaps
- Guidance Reference Table (GRT) approach, to develop specific research questions to guide planned workshops
- Workshops with selected NRAs, to share experience on third-party data
- Synthesis of results, to develop lessons learned and inform guidance for NRA's.

Three specific use cases were selected with Ministry of Infrastructure and Water Management (IenW), Rijkswaterstaat (RWS) and Netherlands National Road Traffic Data Portal (NDW) from the Netherlands and Transport Infrastructure Ireland (TII): Floating Vehicle Data (FVD) for traffic management, ROad MOnitoring 1 and 2 (ROMO1&2) In-Vehicle Data for road and asset management, and eCall Data for road safety and traffic management.

Some of the more significant lessons learned emerging from the deep dive include (see for an elaborate description **chapter 6**):

1. Pre-procurement phase:

- Pre-studies: Pre-procurement initiatives such as proof-of-concepts (POCs), pilots, research and trials are essential for incubating and validating third-party data concepts and aligning objectives before full-scale procurement. For instance, the RWS FVD trial in Zuid-Holland and small-scale POCs for ROMO1 highlighted the necessity of preliminary testing to validate data usefulness and identify implementation goals. Previous Horizon 2020 project on eCall has estimated significant benefits for traffic management, reducing emergency response times by up to 50% in rural areas and 40% in urban areas[1], which improves both traffic flow and road safety. These pre-studies provided valuable insights for further procurement, especially in depicting functional requirements and validating data quality before procurement and full-scale implementation.
- Market consultation: Market engagement informs NRAs of the latest technologies, market trends, and the ever-changing capabilities of third-party providers. IenW and RWS worked on market consultation to tailor ROMO2 procurement strategies and refine data needs for

[1] https://road-safety.transport.ec.europa.eu/european-road-safety-observatory/statistics-and-analysis-archive/esafety/ecall_en

better alignment with technological advancements.

2. Procurement phase:

- Balancing quality and cost: Procurement models, such as the Best Price-Quality Ratio (BPQR), play a crucial role in ensuring that NRAs select both high-quality data and cost-effective solutions. BPQR was used in the RWS FVD and ROMO1 activities which focused on the winning tenders providing the best overall value rather than simply the lowest price. This approach promotes a balance between data quality and cost, preventing the pitfalls of cost-driven procurement.
- Setting minimum requirements and award criteria: Establishing strict data quality criteria during the procurement phase ensures that the procured data meets operational needs. For example, the FVD procurement of Nationaal Dataportaal Wegverkeer (NDW) required real-time data with defined accuracy metrics like Root-Mean-Square Error (RMSE) and Mean Absolute Percentage Error (MAPE), ensuring that low-quality data could not be accepted, even at a lower price.

3. Implementation phase:

- Data integration: Seamless integration of third-party data into existing systems is critical for ensuring real-time accuracy and consistency. The NDW implementation of FVD into its in-house data and the feed to in-house systems like the “Viewer” and “Trajectenmaker” highlighted the importance of robust data integration. The IenW&NDW ROMO1 dashboard, which provides initial visualisation of data before full integration, underscores the need for gradual, verified integration.
- Addressing data attributes: The FVD and ROMO1 use cases emphasized the importance of defining key data attributes like accuracy, coverage, and update frequency during procurement to support smooth implementation. ROMO1’s later realization of additional data needs, such as road temperature, highlights the necessity of clearly specifying technical data attributes upfront. Additionally, mutual understanding of the representation of the data itself (by the NRA as well as the third-party service provider) is of crucial importance.

4. Maintenance and monitoring phase:

- Ongoing data quality assurance and data understanding: Regular monitoring and validation of data quality are essential for maintaining effectiveness of third-party data. NDW’s continuous comparison of real-time data against a ground truth ensured that the data remained reliable for traffic management. Likewise, ROMO1’s focus on monitoring road conditions with third-party data required ongoing checks to maintain the system’s relevance.
- System flexibility: Adaptability in systems and data management processes is crucial for handling updates and changes in third-party data providers. NDW’s in-house systems allowed for in-house integration of updated maps and continuous improvement of data processing methods, ensuring long-term system viability.

5. Overarching phase: Service Level Agreement (SLA), scalability, barriers/challenges, and risk

- SLA monitoring: Effective SLA monitoring ensures that third-party data providers meet the expected standards for data delivery and quality. NDW employed rigorous criteria like timeliness and quality scores to ensure that the data delivered met operational needs. Such criteria were also applied to implementation and maintenance phases for continuous monitoring.
- Scalability challenges: Scaling third-party data solutions to full-scale implementation can last over years and present significant technical challenges. FVD procurement

demonstrated the importance of scalability, with NDW's efforts to increase data coverage while maintaining accuracy and performance.

- Risk mitigation: Addressing potential risks early in the procurement process is vital for ensuring long-term success. The IenW & NDW ROMO2 innovative approach of using variable remuneration as part of the procurement framework mitigates vender locking risk by tying compensation to data quality and volume metric. NRAs should also mitigate risks of a third-party service provider terminating the data stream during the contracting period. TII's workshop highlighted the risk of inaccurate third-party data, such as misinforming stopped vehicle locations, which NRAs must address. The discussions with TII in Ireland also revealed that eCall is an established data source but not yet fully exploited by NRAs as it has been emergency service led. But unlike most third-party data, eCall data is available under European law for road operators to use freely for road safety, without procurement or cost. In Ireland, the data is available from a single national Public Service Answering Point (PSAP) – this may not be the case for all NRAs.

The lessons learned from the five phases highlight the importance of planning, robust procurement models, and continuous data quality assurance. By incorporating these lessons, NRAs can effectively reference, make informed decisions and embark on navigating the complexities of third-party data usage to enhance traffic management, asset management, and road safety across Europe.

What stands out in the lessons learned is that procuring 3rd party data needs a different approach from traditional infrastructure procurement. NRAs should focus on the purpose of the data once retrieved, the goal it is intended for and organisational and procedural impact it will generate. Two items were identified that need further attention (in the next Work Package 5) because of the impact it bears on all phases: which licence does the NRA need to allow use of the 3rd party data as foreseen and which procurement model is best suited to actually acquire the data that is needed?

PRESORT project description

PRESORT is a CEDR Transnational Road Research Programme Call 2022 project aiming to empower NRAs to maximize the potential of third-party data for improved road management. NRAs are increasingly reliant on data to optimize traffic flow, enhance safety, and minimize environmental impact. While NRAs collect some data themselves, a wealth of complementary information resides with data from third-party sources. This data, encompassing information from vehicle manufacturers, navigation companies, emergency services and fleet management providers, holds immense potential for improving road user safety, traffic throughput, and environmental impact. However, NRAs struggle with various challenges, such as data accessibility, quality, governance, and integration with existing systems. PRESORT project aims to address the challenges hindering NRAs from fully utilising this valuable data.

The primary objective of PRESORT is to deliver an *evidence-based decision support guide* that can be used to enable NRAs to make better decisions regarding HOW? and WHEN? to acquire and USE third-party transport data.

PRESORT achieves this with the following approach:

- Capture and understand the current state of third-party data usage by NRAs.
- Analyse the core business needs of NRAs and how third-party data can address them.
- Identify key challenges and best practices in acquiring, using, and maintaining third-party data.
- Deep dive into specific use cases with the highest potential benefit from third-party data.
- Develop an actionable guide for NRAs to effectively identify, acquire, verify, maintain, and utilize third-party data.
- Disseminate the guide to all CEDR NRA member states.

By adopting a collaborative approach with defined roles and responsibilities for each consortium partner, PRESORT follows a sequential work plan with regular engagement with stakeholders, quality assurance reviews, and technical reviews.

The expected outcomes of PRESORT include:

- A comprehensive understanding of the challenges and opportunities related to third-party data usage in NRAs.
- Identification of specific use cases where third-party data can provide the most significant benefits.
- Development of an actionable guide and data catalogue to support informed decision-making in NRAs.
- Dissemination of knowledge and best practices related to third-party data utilization.

By addressing the barriers for accessing and using third-party data, PRESORT aims to empower NRAs to leverage valuable data sources and ultimately enhance their services for the benefit of road users and the environment.

Glossary

ANPR	Automatic Number Plate Recognition
AVG	Actuele Verkeersgegevens - Current Traffic Data
B2G	Business-to-Government
BPQR	Best Price-Quality Rate
CEDR	Conference of European Directors of Roads
C-ITS	Cooperative Intelligent Transport Systems
DRIP	Dynamic Route Information Panel
EC	European Commission
eCall	Emergency Call
EU	Europe Union
FVD	Floating Vehicle data
GDPR	General Data Protection Regulation
GPS	Global Positioning System
GRT	Guidance Reference Table
GRM	Guidance Reference Map
IenW	Ministry of Infrastructure and Water Management, The Netherlands
ITS	Intelligent Transport Systems
MAPE	Mean Absolute Percentage Error
MC	Market Consultation
MEAT	Most Economically Advantageous Tender
NDP	National Data Portal
NCIS	NDW Centraal Informatie Systeem
NDW	Nationaal Dataportaal Wegverkeer. (Netherlands National Road Traffic Data Portal)
NRA	National Road Authority
NTA	National transport authority
OD	Origin-Destination
PSAP	Public Safety Answering Point
QBS	Quality-Based Selection
RIS	Roadside Intelligent Stations
RMSE	Root-Mean-Square Error
RTTI	Real-Time Traffic Information
RWS	Rijkswaterstaat, The Netherlands
SLA	Service Level Agreement
SRTI	Safety-Related Traffic Information

Content

Tables

Figures

1 Rationale and objective

The PRESORT project, a key effort under the CEDR Transnational Road Research Programme Call 2022, is designed to empower NRAs to maximize the potential of third-party data to enhance road management. As the reliance on data-driven decision-making grows, NRAs face challenges in accessing, integrating, and utilising third-party data with their existing systems. PRESORT aims to provide an evidence-based decision support guide to help NRAs navigate these complexities effectively.

Rationale

Leveraging third-party data is a complex, multi-faceted process that offers significant opportunities alongside numerous challenges. While the potential benefits are significant, NRAs must navigate numerous obstacles and risks that arise during the pre-procurement, procurement, implementation, and operation and maintenance phases. These include, but not restricted to, promoting data quality, integrating third-party data with in-house data/systems, monitoring SLA, regulatory compliance, and managing operational complexities and risks associated with these data sets.

Previous PRESORT WP3, revealed substantial gaps between current state and desired state of third-party data usage by NRAs. These gaps highlighted the need for a clearer perception of desirability, technical fit, business viability and other challenges (e.g., visibility, trust, ownership) NRAs face when managing third-party data. To bridge these gaps, WP4 examined real-world practices and experiences, providing guidance to address these issues.

Objectives

The primary objective is to provide an overview of the practices, experiences, and lessons learned from NRAs in procuring, implementing, and utilising third-party data. This report consolidates real-world examples to equip NRAs with practical insights and guidance for effectively navigating the complex landscape of third-party data usage.

Objective 1 (What?): Identify the key considerations in the practices and experiences of NRAs related to the procurement, implementation, and management of third-party data. This includes understanding what the selected NRAs have done (the key considerations) across all phases.

Objective 2 (How?): Examine how NRAs approach third-party data regarding the “what” in objective 1. This involves, to name a few, exploring procurement models, methods, strategies, and tools they employ to overcome operational, technical, and regulatory challenges, including SLA monitoring, scalability, risks and risk mitigation.

Objective 3 (So what?): Synthesise actionable lessons learned from outcomes of the workshops and desktop research. The lessons learned aim to consolidate key insights and transform them into practical recommendations for NRAs. This ensures that the findings from the workshops and desktop research are integrated into actionable recommendation to enable NRAs a quick start to utilise third-party data in traffic management, asset management and road safety.

2 Introduction

The Deep Dive is a crucial component of PRESORT, serving as a key stepping stone toward achieving the project's overall objectives. Building on the outcomes of WP2 and WP3, it focused on an in-depth analysis of three of the four identified use cases, which were selected based on their relevance, feasibility, maturity and potentials to address the rationale and objective outlined in Section 1. The three use cases are:

- Traffic Management using FVD: Enhancing traffic flow and incident management by leveraging real-time data.
- In-Vehicle Data Utilization: Expanding the use of data generated by all types of vehicles for road and asset management, and road safety.
- eCall Data for Road Safety and traffic management improvements: Utilising eCall data to reinforce road safety measures and emergency responses.

The current practices in above areas have already provided tangible lessons and insights that can help NRAs effectively use third-party data. These findings were thoroughly examined through workshops and additional desktop research, forming the basis for practical guidance for NRAs.

The Guidance Reference Table approach adopted established a solid groundwork for NRA guidance to pre-study, procure, and implement third-party data solutions effectively. Moreover, the Guidance Reference Map (GRM) provided structured tables which summarise overall understanding, high-level outcomes, actions and added value from broader lessons learned, for ease of future reference presented in Appendix E.

The structure of this report, D4.1, demonstrates how the deep dive investigation remains focused and strategic, prioritising the three selected use cases. It provides NRAs with referenceable and actionable lessons learned to directly refer to results that could be applicable to their specific needs and contexts in the future:

Section 1 Rationale objectives and scope - Defines the rationale and objectives of D4.1.

Section 2 Introduction - Outlines WP4's role in PRESORT project, introduces the selected use cases, and explains the approach taken in the workshops.

Section 3 Methodology - Details the methodological framework used, focusing on the workshops and the GRT building blocks.

Section 4 Deep dive workshops - Provides background and key findings emerging from the workshop conducted with two NRAs.

Section 5 Data stream exploration - Presents the data requirements and criteria in procurement, among which many are used in implementation, maintenance and monitoring phases. Moreover, it presents an empirical data analysis of the use cases data attributes.

Section 6 Main lessons learned - Summarises the main lessons learned, and the detailed outcomes and lessons learned are presented in GRM tables for user-friendly references in Appendix E.

3 Methodology

To address the "what," "how," and "so what" of the challenges NRAs face in procuring and using third-party data, the methodology (see Figure 1), illustrates the four building blocks used: **Desktop research** (Section 3.1), the **GRT approach** (Section 3.2), **Workshops** (Section 3.3), and **Synthesis** (Section 3.4) and the logical flow to meet the WP4 objectives.

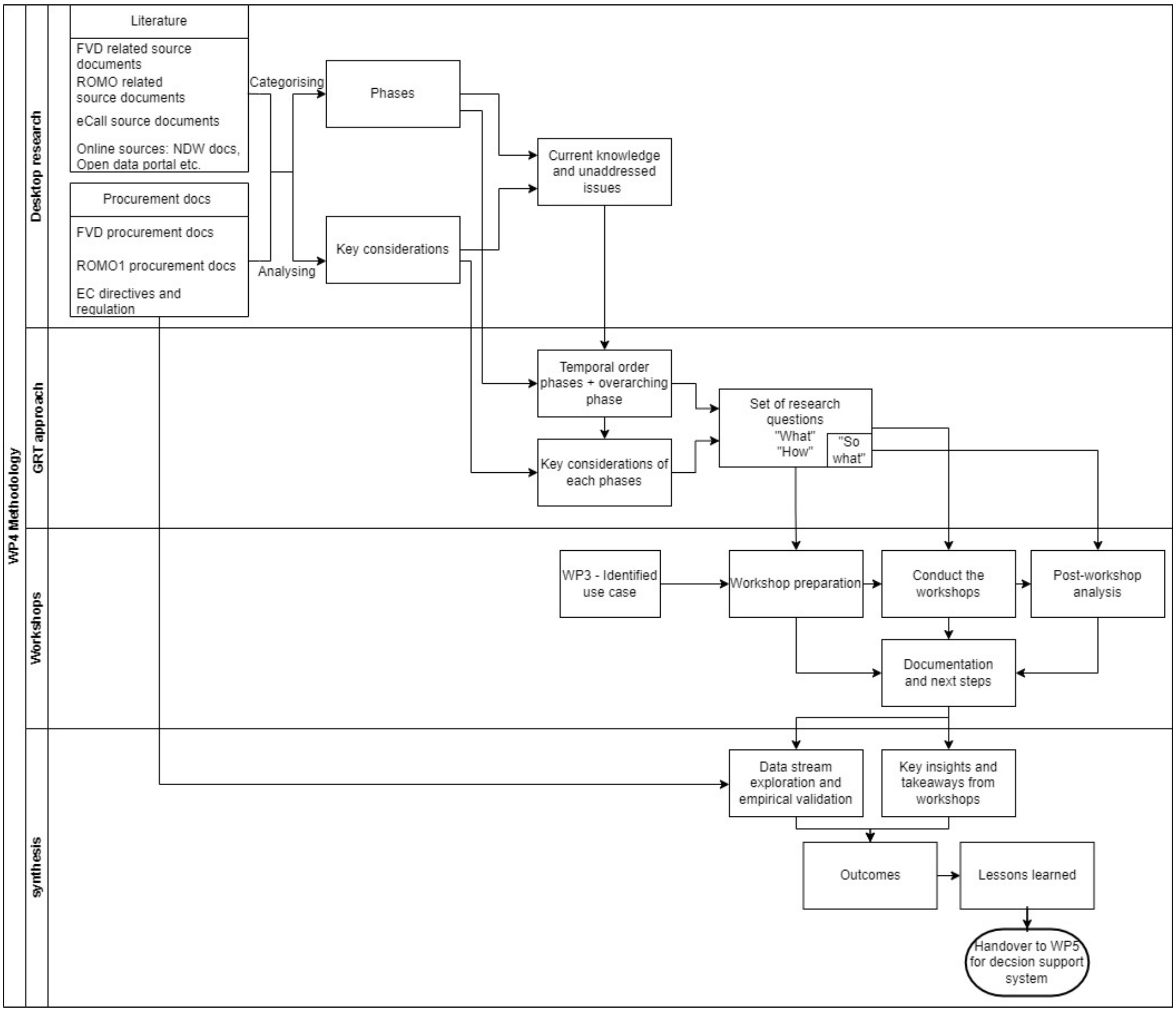

Figure 1 WP4 methodology flowchart

3.1 Desktop research

Baseline information was gathered related to FVD, ROMO1&2, eCall, along with relevant regulatory EC documents, procurement documents from NRAs and other sources. This was used to inform the subsequent GRT approach, workshops and Synthesis. Through examination, gaps were identified and key considerations framed for each of the phases in the third-party data procurement and implementation process.

Literature documents

Building on the literature insights from WP2, the literature review focused on exploring various areas of domain knowledge:

- FVD-related source documents: These include technical literature and reports on

Floating Vehicle Data, outlining the system's functionalities, use cases, and the challenges experienced by various NRAs during the deployment of FVD.

- ROMO-related source documents: ROMO1 source materials offer insights into the system's development, procurement processes, and how data is used for road management purposes.
- eCall-related source documents: These materials detail the deployment of eCall systems and their potential as a source of real-time traffic data for road safety and incident management.

Additionally, online sources such as NDW docs, Open Data Portal, and other public records were explored to provide context and additional clarification for the NRAs' experiences with data procurement and usage.

Procurement documents

Procurement documents are an essential consideration which helped clarify the legal, technical, and financial considerations NRAs faced when working with third-party data:

- FVD procurement documents: These provide specific details about the acquisition of FVD services, including SLAs, contractual obligations, and performance indicators.
- ROMO1 procurement documents: These documents discuss the procurement strategies for the ROMO1 system, including the roles of vendors and the expectations set in the procurement agreements.
- European Commission (EC) procurement documents: EC procurement documents offer insights into the broader regulatory framework of public procurement.

Some of the FVD and ROMO1 procurement documents were mentioned during the workshops and were shared by participating NRAs after the workshops. These documents provided essential information about the third-party data procurement complimenting the workshop results. Some of the (pre)-procurement documents were discovered by self-search beforehand that provide complementary support of understanding the NRAs procurement documents.

Delineate the five phases and key considerations

After collecting and reviewing abovementioned documents, we analysed them and identified gaps in knowledge — areas where the existing knowledge fails to provide sufficient answers. This process involved:

1. Define the five phases, which is further discussed in Section 3.2:

- Pre-procurement phase
- Procurement phase
- Implementation phase
- Maintenance and monitoring phase
- SLA, scalability, barriers/challenges, and risks

2. Key considerations: For each of the above phases, the desktop research identified key considerations that needed to be addressed. These included:

- Pre-procurement: Pre-studies to set procurement strategies and establish metrics.
- Procurement: Rationale, data requirements, procurement model, licensing etc.
- Implementation: Integrating into existing data and systems, system operating etc.
- Maintenance and monitoring: Ongoing performance evaluation, system upgrade etc.
- SLA, scalability, barriers/challenges, and risks

Through desktop research, we gained current knowledge and the unaddressed issues about the practices. This groundwork helped structure the set of research questions for the workshops, where they were explored in greater depth with the participating NRAs.

3.2 Guidance Reference Table approach

The **GRT approach** was used to define the set of research questions that guide the workshops. This approach systematically formulated knowledge from desktop research and proposed specific questions on the process of NRAs procuring, implementing, and maintaining third-party data sets. It operated on two dimensions:

1. Delineating the **phases** (distinctive temporal order phases and overarching phase) involved in NRAs' engagement with third-party data.
2. Identifying the **key considerations** within each phase, forming the basis of the research questions guiding the workshops. Table 1, below, presents these questions, each research question (in numerical order) addressing a key consideration. Due to space limitations, the questions are abbreviated, and ad-hoc questions raised during discussions are omitted. The questions begin with the 'what,' followed by the 'how' and 'so what' when the 'what' answers warrant further exploration.

The key considerations were used to formulate a set of research questions. In Appendix E, the detailed outcomes were organised in five tables using the GRM format, corresponding to the five phases. As the workshops addressed these research questions, the key considerations from the GRT approach evolved into consolidated key considerations.

In the original plan [2], Task 4.2 of WP4 described items to be discussed during the workshops. Three sequential phases: procurement, implementation, and maintenance and monitoring were originally proposed. By using the GRT approach, we expanded this to include the pre-procurement phase; and overarching phase including the SLA, scalability, barriers/challenges and risk, which either suggest future considerations or cut across multiple phases of NRAs' engagement with third-party data.

[2] Part A PRESORT AECOM - CEDR Call 2022 - Data.pdf

Table 1 Defining research questions using Guidance Reference Table approach

Research question	Pre-procurement phase	Procurement phase	Implementation phase	Maintenance and monitoring phase	SLA, scalability, barriers/challenges and risk
1	What key considerations were used to define/select the specific goals FVD/ROMO/eCall data should achieve?				
2		What key considerations was followed to identify the most useful data sets?			
2a		How do the data sets meet these key considerations and how were they used to evaluate?			
2b		Were there any specific challenges in selecting the data sets, and how were they addressed?			
3		What data quality requirements (incl. transition to the next successor) were defined and how were they chosen?			
4			What processes were put in place to integrate third-party data into the existing data/systems? How were these processes managed and implemented?		
4a			What were the major		

			milestones and products in this phase, and how were they achieved?		
4b			What are the technical or organizational challenges, risks encountered?		
5				What key considerations were used for this phase?	
5a				How was the maintenance phase executed to ensure the continued data quality and reliability?	
5b				What ongoing processes are in place to monitor and maintain the operation? how?	
6					How are the SLA being ensured in and beyond the contract period?
7					What strategies were employed to the scalability key considerations in present and in the future?
8					What barriers/challenges and (potential) risks were encountered and how are they mitigated in the above four phases? What are the insights and key takeaways?

3.3 Workshops

The **Workshops** explored the experiences, practices, and challenges of selected NRAs with the aim to use the outcomes to develop use case-oriented lessons for all NRAs.

The interaction, shared insights, and collaborative process considered complex issues related to third-party data usage by NRAs, and the planned structure of the workshops enabled real-time feedback and shared experience, which investigated existing practices, validated and refined desktop research, and ensured that the practical experience and lessons learned aligned with (their) stakeholder needs.

Each workshop consisted of four steps:

1. Workshop preparation

- Select NRAs and use cases: Choose NRAs that are diverse in terms of geography, third-party data usage maturity level, governance models, and challenges faced. The diversity will help extend workshop findings that can be generalized and guidance that can be adapted to different contexts.

The selection of NRAs from the Netherlands and Ireland was based on their strategic importance and the relevance of their existing infrastructure.

- The Netherlands was chosen due to its advanced Intelligent Transport Systems (ITS) initiatives and its leadership in integrating FVD for traffic management [1] [2]. It has been proactive in exploring innovative data-driven approaches, making it an ideal candidate for studying third-party data usage in enhancing traffic efficiency and road safety.
- On top of trialling with third-party data, TII was selected due to its focus on using new technologies for road safety improvements, particularly exploring use of eCall data. TII involvement in initiatives aimed at improving emergency response times via data collected from eCall data aligns with the PRESORT objectives.

Use case qualification

Our desktop research examined the background of various use cases identified in WP3. Based on the qualification criteria outlined in Table 2, the FVD, ROMO, and eCall use cases were selected, corresponding to traffic management, asset management and road safety.

Table 2 Use case qualification matrix

Criteria	FVD TM (NL)	Road Monitoring (NL)	eCall Data Use (IE)
Data Availability	High	High	Medium
Technical Feasibility	High	High	Medium
Business Viability	High	Medium	High
Impact on Road Safety	Medium	High	High
Ease of Implementation	Medium	Medium	Medium
Current practice of data utilisation	High	High	Low
Procurement maturity level	High	High	N.A.
Stakeholder Interest	High	High	High

The qualification criteria considered were:

- **FVD (NL)**: The high availability of Floating Vehicle Data and the strong technical infrastructure in the Netherlands, which supports its integration provided rich potential to deep dive. The procurement maturity level and the business viability are high, given the potential for improving traffic flow and reducing congestion, making it a key focus area for the NRA.
- **Road Monitoring (NL)**: The ROMO use case leverages data from third-party in-vehicles to monitor road conditions, offering high impact on road safety. The data is procured and available, and the technical feasibility is well-supported by existing systems from the procurement. The stakeholder interest is also high, making it an attractive use case for further investigation.
- **Third-party data and eCall Data Use (IE)**: Selected for its high relevance to road safety improvements and traffic management, whilst acknowledging the early maturity of eCall data exploitation, the potential impact on emergency and road operator response times and road user safety is significant. Procurement is not relevant since any NRA in Europe can access eCall data at no cost, often from other government departments, however there is huge potential and cost associated with its exploitation. The business viability and stakeholder interest are high, justifying its inclusion in the detailed workshop investigation.

- Develop workshop setups and agendas: Agendas and process of engagement during each workshop, focused on:
  1. Presenting the background and introducing each use case provided by NRAs;
  2. Meeting the workshop objectives by addressing the research questions and targeting key considerations across the four phases;
  3. Discussing challenges and potential risks in the overarching phase.

2. Conduct the workshops

- Introduction of PRESORT project and use case background: Each workshop began with an overview of PRESORT project's background, covering PRESORT objectives, rationale, and the specific objectives along with the set of research questions for the workshop. Representatives from FVD, ROMO, and eCall provided brief presentations on the background and current practices related to their respective use cases.
- Breakout parallel sessions: For the workshops with the NL NRA, two parallel sessions were held to facilitate efficient discussions to answer the set of research questions on traffic management and asset management. In the workshop with the IE NRA, the focus was divided into two parts: general third-party data and eCall data, both discussed within a compact one-hour workshop.
- Reporting back and preliminary conclusions: After the breakout sessions, participants reconvened to share insights, raise additional questions, and summarised immediate findings.

3. Post-workshop analysis

- Refinement of meeting transcripts: each workshop was recorded with the consent of participants. Post-workshop, the transcripts were refined to check clarity, conciseness, and removal of irrelevant or faulty information.
- Extraction of insight and takeaways: The transcripts were comprehensive and verbatim. Key insights and takeaways were extracted through analysis and presented in table format, providing a clear and concise overview of the workshops without losing focus on core issues.

4. Documentation and next steps

- Appendix A and B workshop reports: These appendices documented the key discussions and findings from each workshop, serving as a reference for Section 5 and 6.

- Feedback loop: The documentation was shared with participating NRAs to establish a feedback loop, promoting ongoing communication and refinement of the outcomes and lessons learned as the project progresses.

3.4 Synthesis

Synthesis integrated the outcomes from workshops and desktop research to form lessons learned. By unifying findings from all methodology building blocks, synthesis consolidated the knowledge gathered and transformed them into actionable lessons learned for the NRAs.

The input for synthesis comes from two main sources:

1. Key insights and takeaways from workshop analysis:

The insights were consolidated and further refined with support from desktop research, which provides additional context and depth to the workshop outcomes. The outcomes drawn were supported by both real-world practices and desktop research knowledge.

2. Data stream exploration and empirical validation:

During the workshops, NRAs also shared data validation methods that they used to explore and ensure the integrity of their data streams. These techniques, combined with desktop research, helped justify the accuracy and applicability of the data streams. Synthesising this information allowed for a solution-based understanding of how third-party data can be effectively validated, integrated, and scaled within the NRAs' systems.

4 Deep dive workshops

The three deep dive workshops explored many different facets of the specified use cases with relevant stakeholders. The introduction of the workshops use cases were described below, with logs of the detailed discussions answering the set of research questions provided in Appendices:

4.1 Traffic management and asset management introduction – Netherlands

Background

The NDW in the Netherlands serves as the National Data Portal for traffic data from various road authorities. Its primary objectives are the procurement, collection, and distribution of high-quality traffic data to support traffic management, information services, and policy analyses.

NDW represents a unique collaboration between multiple stakeholders, including the Ministry of Transport, Public Works and Water Management (IenW), provinces (such as Noord-Holland, Zuid-Holland, Noord-Brabant, and Utrecht, etc.), municipalities (such as Amsterdam, Utrecht, and Rotterdam), and city regions.

While NDW collects data from numerous sources, including Roadside Intelligent Stations (RIS), this in-house data is not sufficient to cover the entire targeted network. At the start in 2008, NDW aimed to cover a basic network of around 6600 kilometres of motorway, provincial, and urban roads within 3 years [3]. To achieve this goal, NDW began procuring third-party data from external traffic data providers, following a formal competition process.

The NDW provides minute-by-minute data on various aspects, including traffic speeds, intensities, travel times, road works, bridge and lane statuses, road blockages, events, control strategies, incidents, and traffic management messages. This comprehensive dataset is made available to service providers, traffic authorities, and other stakeholders to improve traffic management and information services across the Netherlands.

4.1.1 Floating Vehicle Data introduction

Currently, NDW has access to many roadside equipment (loop detectors, cameras etc.), but not enough for a nationwide coverage. In addition, although loop detection of vehicle data is with almost no latency, but the detected values are less stable and not flexible nor scalable, due to the expensive implementation and maintenance cost.

The **rationale** for procuring FVD is to increase the travel time coverage and minimise the data collection cost. By doing so, FVD are expected to enhance real-time traffic monitoring, enabling quicker detection of congestion and incidents. It improves traffic management by optimizing flow and offers comprehensive data coverage, even on smaller roads not covered by traditional sensors. FVD also supports predictive analytics for long-term traffic planning. It is studied and proven to be more cost-effective than installing and maintaining physical traffic sensors

The Netherlands has been at the forefront of utilising FVD data for traffic management and real-time traffic monitoring. In Zuid-Holland, a pioneering pilot project was initiated, which incorporated Automatic Number Plate Recognition (ANPR) data into the traffic management system to enhance real-time traffic monitoring and optimize the accuracy of travel time calculations by replacing ANPR data with FVD. The pilot was a success and laid the groundwork for the broader application of data-driven traffic management across the Netherlands, contributing valuable insights to subsequent procurement processes.

Over the past few years, NDW has conducted three major procurement processes for FVD, each building upon the success and lessons learned from the previous ones. The scope of data collection and application has expanded, providing NDW with increasingly comprehensive and sophisticated data solutions for improving road network operations across the Netherlands. The first two were both won by Be-Mobile. At the time of the report, the third tender for FVD was also awarded to Be-Mobile, a key third-party data provider in the NDW's efforts to integrate real-time vehicle data into national traffic management systems.

The **objective** is to acquire off-the-shelf third-party FVD data which NDW can use for further data aggregation. In this Request for Proposals, real-time FVD speed data on unit segment of the whole Netherlands road network is requested by NDW. NDW will aggregate the off-the-shelf third-party FVD data (combining with predefined trajectories) to calculate metrics such as speed, travel times or even for developing other NDW products, for usage within the government organizations affiliated to NDW. Quoting from the third tender [4], NDW requests the following:

For this, NDW requests the following delivery:

- Real-time speed data based on FVD
  - On all available roads within the geographical area of the Netherlands including connecting (main) roads up to at least 10 km across the borders with Germany and Belgium.
  - With an availability of 24 hours, 7 days a week all days of the year.
  - With an update frequency of 1 minute.
  - If available in Open Location Referencing format (OpenLR)
  - If available, the map as shape / geojson
  - In the smallest available segments.
  - Including information on the length of each segment.
  - Through a http(s) pull interface on predetermined moments.

NDW will use this data for:

- Calculation of travel times on routes based on the delivered speed information in segments.
- Real-time delivery of data to governmental organization linked to NDW for traffic management applications (use within Traffic Management Centres, operation of DRIPs, travel time predictions etc.) among others uses.
- Historical storage.
- Use in historical analyses.

The above uses are limited to (clients of) governmental organizations. Government organizations may, however, have unlimited access to the data. The contract duration is 2 years data delivery (calculated from the start of full data delivery) with the option of 2 extensions of 1 year. NDW requests third-party data providers to indicate a price on the following option: **distribution of aggregated data to all parties outside government**. This means that data processed by NDW (e.g. travel times over a complete road section) is distributed as open data with no restrictions on use. It is mandatory to give a price for this option. With this statement in the procurement document, NDW embeds the key consideration of data licensing and usage in the procurement phase.

Figure 2 Attendance of FVD workshop (full) and ROMO workshop (partial)

4.1.2 Road monitoring introduction

Background

The ITS Directive[3] [5], mandates that safety-related minimum universal traffic information (SRTI) should be made available to users free of charge, where possible. With that incentive, NL NRA expends to procure data for road management purposes for a win-win situation between NRA and third-party data provider, in this case, the car manufacture.

Rationale

Traditional asset management methods, such as periodic measurements using complex infrastructure systems are costly. NL NRA notices more budget constraints to develop physical infrastructure, while road management coverage needs to be increased. Therefore, they look for innovative/alternative solutions in asset management.

The **objective** is to see the validity of using probe vehicle data for asset management, where in-vehicle data is used to create services for road authorities that help to gain overview into damage, missing street furniture and unsafe situations.

Throughout the years, there were four to five small-scale POCs, for example, in Noord-Brabant, to investigate the benefits in probe vehicle in-vehicle data. Rijkswaterstaat (RWS), SmartwayZ.NL, and the Ministry of Infrastructure and Water Management (IenW) have conducted successful trials with this technology. Based on these positive outcomes, the decision has been made to upscale the initiative nationwide [6].

In the pre-procurement phase, aiming for a flexible goal setting scheme with such an innovative solution, functional requirements and descriptive use cases were set instead of quantifiable criteria was set to validate whether the procured data met the goal. For ROMO1, ten use cases were identified in three categories: Asset management, Winter management, and Road safety. ROMO1 focuses on utilising factory-installed sensors in modern vehicles to gain new insights into critical aspects of road management. This approach represents a shift from

[3] Delegated Regulation (EU) No 886/2013

Conférence Européenne des Directeurs des Routes
Conference of European Directors of Roads

traditional methods of road condition assessment to an innovative solution. By leveraging vehicle sensor data, the project aims to complement and potentially enhance traditional observation methods, providing a more comprehensive and up-to-date view of road conditions across the Netherlands. ROMO1 project is not about the procurement of raw data, but rather to obtain services that display on a dashboard to provide road managers with actionable information about:

- Road surface conditions
- Engineering structures
- Other related aspects of road infrastructure

This information will be presented on a dashboard, offering road managers valuable insights for more effective and efficient management of the road network [6].

The NDW is the purchasing body for its partners. On behalf of the State of the Netherlands, NDW is initiating a procurement process of the Road Monitor (ROMO) project. This procurement process is in accordance with the open procedure on the instructions of its partners, IenW and Rijkswaterstaat.

Market consultations

In 2023, a market consultation was issued by the IenW in preparation of the initiative ROMO2 (a follow-up to ROMO1), to communicate with the market or relevant experts and to gain more information and understanding of what is available on the market. As part of this market consultation, the IenW was seeking information from technology providers that offered turnkey solutions and had a proven track record of providing services to road authorities. ROMO2 wishes to use several service providers simultaneously and focuses on the applicability of the service provided to the daily practices of the road authorities. With the market consultation, IenW requested to gain insights on[4]

1. Current Business-to-Government (B2G) product portfolio’s
2. The appetite of industry players to act within a coopetition ecosystem
3. The soft cooperation rules under which an individual player wants to cooperate
4. The constraints and boundaries an individual player might have
5. Suggestions on the organisation of such an ecosystem
6. Suggestions on the financial mechanism of such an ecosystem
7. The willingness to, if any, further develop customized data products with NL NRA

Figure 3 Attendance of ROMO1 Workshop Participants (partial)

[4] https://www.tenderned.nl/aankondigingen/overzicht/298865

4.2 eCall data and third-party data introduction – Ireland

Road safety is a critical concern in the EU, and **eCall**, an automated emergency call system, plays a key role in reducing fatalities from road collisions. The system automatically notifies emergency services after a crash, allowing faster response times and potentially saving lives. Since March 31, 2018, eCall has been mandatory in all new cars and light commercial vehicles across the EU, with up to 20 million new vehicles integrating this feature annually [7][8]. The European Commission monitors eCall's impact as part of its broader efforts to improve post-crash care and achieve **Vision Zero**, the goal of eliminating road deaths by 2050 [9].

The eCall system offers numerous benefits, particularly in emergency response and traffic management. When activated by airbag activation or manually by vehicle occupants, eCall automatically sends vital data (such as the vehicle's location, make, and fuel type) to emergency services via cellular network, reducing response times and increasing the chances of saving lives. This system is especially valuable on rural roads or areas with limited Intelligent Transportation Systems (ITS) equipment, where quick location of incidents could significantly reduce the severity of accidents and minimize traffic disruptions.

The potential for eCall to improve road management lies in its ability to enhance real-time incident detection and response, providing authorities with immediate data on vehicle accidents or breakdowns. This information helps reduce secondary accidents and traffic congestion. Additionally, it could be integrated with traffic management systems to further streamline the coordination of emergency services, improving overall road safety and operational efficiency.[5]

Although eCall’s primary function is emergency response, its integration into NRA’s road management systems could address critical gaps, especially on roads lacking ITS equipment. By processing and filtering out false alarms, as done by systems like TeCall[6], eCall data can help build a business case for its broader use. The timely and accurate information (particularly in airbag activation) on vehicle stops, crashes, or other incidents allows for quicker decision-making by road operators, potentially preventing further congestion or hazards.

This integration would not only enhance road safety but also contribute to more efficient use of resources, particularly in situations where rapid responses are crucial. Gathered from the eCall workshop with IE NRA, it is a long-term goal to prove eCall’s value in traffic management, along with ensuring that the eCall data (owned by the emergency services) can be fully exploited for optimizing road network performance.

The discussions with TII in Ireland showed that eCall is an established data source but not yet fully exploited by NRAs as it has been emergency service led. But unlike most third-party data, eCall data is available under Eu law for road operators to use freely for road safety, without procurement or cost. In Ireland the data is available from a single national Public Service Answering Point – this may not be the case for all NRAs

Ireland is currently exploring the use of the eCall data through a series of proof of concepts and establishing the sustainability of feeds to TII long term.

[5] https://www.intertraffic.com/news/traffic-management/ecall-traffic-management
[6] https://vesos.co.uk/article/vesos-announces-tecall-concept

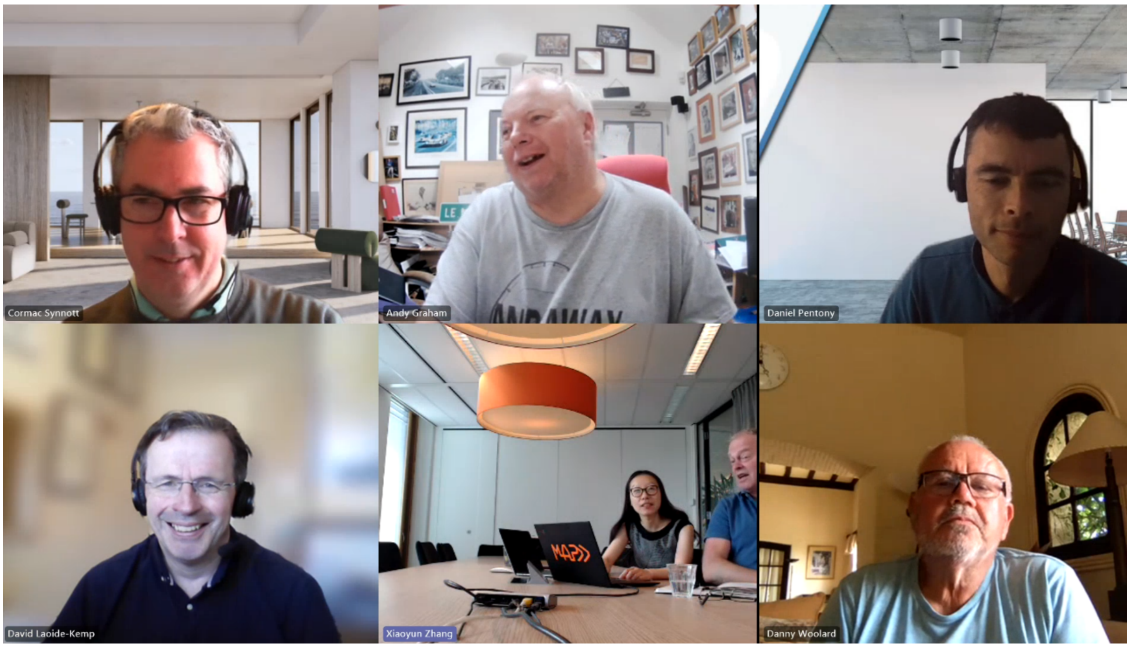

Figure 4 Attendance of eCall Workshop Participants

4.3 Workshop log

The workshop logs are detailed in Appendix A: Report on workshops with the Netherlands NRA and Appendix B: Report on workshop with the Ireland NRA. These logs are synthesised from the full transcripts of the workshops. Rather than providing the entire transcripts, the key insights and takeaways, which address the research questions, are organized according to the workshop agenda. These logs form the foundation for the in-depth data analysis in Section 5 and the main lessons learned presented in Section 6.

Disclaimer

The views, comments, practices, and experiences shared during these workshops are provided for informational purposes only to PRESORT, a CEDR research project, and should not be construed as official statements or positions of the participants' respective organisations. Participants are encouraged to critically evaluate the outcomes, lessons learned and challenges of NRA procuring and using third-party data. PRESORT reserves the right to interpret and utilize the information discussed during the workshops. While efforts are made to ensure the accuracy of shared content, the information presented may not always be fully accurate, complete, or up to date. Neither PRESORT nor the participants are liable for any inaccuracies or misinterpretations of the shared content.

5 Data stream exploration

This section presents the results of Task 4.3, which stated to analyse two third-party data sets procured by NRAs. FVD and ROMO1 were selected due to their successful procurement processes, advanced data requirements and criteria, and high relevance to NRAs in road and traffic management. However, since the data from FVD and ROMO1 was considered commercially sensitive, it was not able to be shared for direct analysis. Instead, we conducted desktop research to study the procurement procedures of FVD and referenced the ROMO1 publication in Appendix D: ROMO1 data paper.

In Section 5.1, we examine the data requirements and criteria used in the procurement, including procurement models, tender methods, minimum requirements, and award criteria. Section 5.2 describes the attributes of the two data streams as detailed as possible, aligning them empirically with the established requirements and criteria. By doing so, we gain more understanding of the consolidated key considerations in the procurement phase, such as data requirements, criteria, licensing and usage, as well as in the implementation phase, such as covering data attributes specifications, data integration, system operations, data extraction, and data access control.

5.1 Data requirements and criteria

The European Commission’s public procurement strategy focuses on six strategic policy priorities to improve EU public procurement practices in a collaborative manner by working with public authorities and other stakeholders.[7] NRAs procurement of third-party data needs to be aligned with European public procurement directives 2014/23/EU, 2014/24/EU and 2014/25/EU [10][11][12], particularly to Article 82: Contract award criteria. The EC also supports national authorities with large infrastructure projects through a voluntary ex-ante assessment of the procurement aspects. Governments in the European Union must follow a tendering procedure in accordance with the directive. In the Netherlands, these rules are included in the Public Procurement Act 2012[8][9].

Rooted from EU procurement, Section 5.1.1 presents several relevant procurement award models and general introduction of the ones applied to the workshops. Moreover, this section presents the tender methods of data requirements and criteria mentioned in deep dive workshops.

This section outlines the specific data requirements and criteria during the procurement phase of third-party data. It sets the groundwork for WP5 to generalise the data requirements for the decision support system guidelines, offering insights to other NRAs on how to evaluate third-party data and its providers.

5.1.1 Procurement award models

NRAs and third-party data providers often use award models that balance price and quality when procuring services or awarding contracts. Discussions with the NRA’s identified the following:

1. MEAT (Most Economically Advantageous Tender) is an older term, now largely replaced by Best Price-Quality Ratio (BPQR), but still used in some contexts. It also aims to

[7] https://single-market-economy.ec.europa.eu/single-market/public-procurement_en
[8] https://wetten.overheid.nl/BWBR0032203/2022-03-02/0
[9] On 1 July 2016 the amended Dutch Public Procurement Act entered into force. The amended Public Procurement Act implements the latest EU procurement directives (2014/23/EU, 2014/24/EU and 2014/25/EU).

select the tender that offers the best Value for Money (VfM), considering factors beyond just price. MEAT model considers technical merit, functional characteristics, and quality alongside cost, which aims to foster competition and ensure transparency. MEAT is guided by EU directive 2014/25/EU [11] and national laws in public procurement.

2. Best Price-Quality Ratio

BPQR is an award model that aims to identify the tender offering the best VfM. It considers both cost and quality factors to determine the most advantageous offer. BPQR replaces the earlier MEAT model, focusing on balancing scope, quality and price. It uses a scoring system with weighted criteria, which can include qualitative, environmental, and social aspects. This model is further explained in Section 5.1.2 with results retrieved from the FVD workshop.

Within the broader overarching principle of VfM in procurement, the BPQR award model is applied to achieve VfM. To this end, the BPQR balances between price scores, scope (which may include licensing the use of data to other parties, e.g. governmental institutions or municipalities) and quality using the Price-Quality Ratio. The balance is self-defined by NRA and the bid with the best Price-Quality Ratio is awarded.

In the most recent ROMO2 procurement[10], the quality of VfM takes 80 out of 100, whereas the cost takes the other 20 out of 100.

Desktop research showed the other commonly used award models, which include:

- Lowest Price, which simply awards the contract to the lowest bidder. While straightforward, it's often criticized for potentially sacrificing quality.
- Quality-Based Selection (QBS), which focuses primarily on the quality and competence of the bidder, with price negotiations occurring after selection.
- Value-Based Procurement, which considers the total value created over the lifecycle of the product or service, including factors like innovation and sustainability.

3. Procurement framework

A procurement framework is quite often used as a structured approach to managing third-party data procurement. The main character of procurement framework is that the payment or rewards are linked to performance metrics or outcomes. It incentivizes third-party data providers by tying compensation to the achievement of specific goals, such as data volume, data quality, cost savings, timely delivery or efficiency improvements etc.

A conceptive and innovative approach to procurement is to develop a procurement framework, which was mentioned in ROMO workshop with NL NRA. This framework adopts a variable remuneration scheme in procurement. And it adopts a volume quality matrix to describe a flexible procurement framework. More details can be found in Table 11, it is mostly general information as it is still under conception.

Since procurement award models play such a pivotal role in the selection of the 3rd party data provider and the (quality of) data itself, this topic will be given additional attention in the next Work Package 5.

5.1.2 Tender methods

The main tender methods that an NRA uses for procuring third-party data are: Open procedure, Restricted procedure, Competitive dialogue and so on. FVD and ROMO1&2 procurements chose an Open procedure[11] [12] to encourage competitiveness. This section

[10] https://www.tenderned.nl/aankondigingen/overzicht/352683

[11] Public procurement procedures via a notice on TenderNed https://www.tenderned.nl/cms/nl

[12] Public procurement procedures via a notice on Tenders Electronic Daily (TED Europe https://ted.europa.eu/en/

presents the details of the tender methods and evaluation methods applied in the procurement phase of third-party FVD data.

The FVD procurements by NDW have been using this tender method for all three of the procurements. NDW evaluated bids based on quality and assigned a score to them. The total quality score is then raised to the power of 3 and divided by the price, providing insight into which offer provides the most quality per euro.

Score = (Total quality score) ^{3} / price

Tender planning procedure

A clear tender procedure timeline was presented, including several phases where each phase consists of several milestones.

Table 3 Example tender procedure timeline of the third FVD procurement

Phases	Tender Planning Milestones	Dates (only examples)
Pre-tendering	Sending of the request for proposal	xx May 2024
	Question round: Deadline for submitting questions and comments	xx May 2024
	Written response to questions in the Information Memorandum	xx May 2024
Tendering and decision-making	Final date for submitting the proposal	xx June 2024
	Start of delivery of real-time test feed (14 days)	xx June 2024
	Provisional award and invitation to the highest-scoring participant on EMVI criteria	xx July 2024
	End of objection period	xx July 2024
Implementation phase	Start of implementation and testing phase	xx October 2024
Operational phase	Start of operational phase (operational data delivery)	xx April 2025

During the tendering and decision-making phase in Table 3, NDW evaluated the submitted offers based on the criteria detailed in minimum requirements and award criteria.

Minimum requirements

The minimum requirements set the baseline thresholds that a bid must meet to qualify. This ensures that bids offering lower data quality, even if accompanied by an attractive price, cannot succeed if they do not meet the minimum requirements [4]. This ensures that only providers that can provide data above a certain threshold will be taken into consideration during the remainder of the bidding phase, thus limiting the amount of work to be executed during procurement.

Table 4 Minimum requirement in the FVD tendering by NL NRA

	Minimum requirement	Element details
M1	Agreement with assignment specifications	- Guidelines of live real-time data feeds
M2	Minimum coverage	- Areas, segments - Number of probes relative to the total traffic - Map display - Comparison with chosen NDW inhouse 125

		measurement points of a fixed 2-week period
M3	Minimum completeness	- Completeness of data is assessed by comparing regional traffic management network, detailed in [13]

Award criteria

All bids that met the minimum requirements were evaluated by NDW evaluation team based on the award criteria. The evaluation team consisted of at least three people, including at least one representative from the partners. The bid amount and any option prices submitted with G1 were not disclosed to the evaluation team. Each member of the evaluation team first assigned a rating individually. During a plenary evaluation session, the evaluation team discussed the individually assigned ratings. The goal of this plenary session was to reach a consensus on the rating to be assigned.

Table 5 Award criteria in the FVD tendering by NL NRA

No.	Criterion	Conditions	Methods
G1	Bid amount	Maximum budget	Non-disclosure of actual budget amount
G2	Quality of traffic data	N/A	
G2.1	Availability of data	Minimum score 200	Score system from 200 to 1000
G2.2	Timeliness score	Minimum score >0	Score system from 0 to 1000
G2.3	QKZ method	Minimum score >0	Score system from 0 to 1000
G2.4	Quality score MAPE & RMSE	Minimum score >0	RMSE value: Score system from 0 to 500 MAPE value: Score system from 0 to 1000
G2.5	Implementation, management and maintenance	Minimum score 200	Score system from 200 to 1000
G2.6	Filtering and quality assurance	Minimum score 200	Score system from 200 to 1000

5.1.3 Open Knowledge Bank

NDW represents NL NRA to collect, procure, aggregate and provide various types of traffic data, including but not limited to the following, which is reworked and published on the open data portal[13].

- Traffic intensity: the number of vehicles passing a measuring point.
- Local speed: average speed of vehicles passing a measuring point.
- Realized or estimated travel time: based on in-house data and procured third-party data, such as FVD.

[13] https://opendata.ndw.nu/

- Vehicle category: derived from the length of passing vehicles.

As part of the Netherlands' NRA, NDW has established itself as the centre of excellence for data management and innovation. With teams of skilled data scientists, developers, and technical experts, NDW has built a robust knowledge base and infrastructure, positioning it as the NDP for traffic and mobility data in the Netherlands. This central role enables NDW to offer essential data services, from real-time traffic information to advanced analytics, supporting both the public and private sectors in traffic management, policy development, and innovation. Additionally, NDW provides a documentation portal[14] to support the interpretation of these various traffic information products and services, which is freely accessible to the public. The tender methods mentioned in Section 5.1.2 are also available and detailed on this portal.

5.2 Deep dive on two example data streams

The deep dive of two third-party data streams is to review and explore the reliability, accuracy, and utility of data procured by NRAs. In this section, we present a conceptual validation of two data streams — one focusing on FVD and the other on ROMO data. The validation process, due to data-sharing constraints, is adapted to conceptual validation, which met the non-sharing procured raw data to non-governmental organization rules from these two NRAs.

The FVD stream, provided real-time traffic information based on Global Positioning System (GPS) signals from moving probe vehicles. Conceptually, the procured FVD stream was validated based on the criteria outlined by the NDW tender methods in Section 5.1.2. The validation process involved simulating a real-time environment where the procured FVD data stream (over a specified period, typically two weeks) was analysed, to check consistency and accuracy across different traffic conditions. Given the challenges of directly sharing NRA procured third-party data, the validation of such a data stream was out of the scope.

Thus, we only presented the empirical methods of the data stream validation based on the information we gathered from the workshop. The validation of the FVD data stream involves comparing the real-time data against the predefined minimum requirements and award criteria, as established in the NDW tender process. The key parameters under scrutiny include [14]:

- Data coverage and completeness: This entails assessing whether the data adequately covers the specified geographic regions and whether the data points are sufficiently dense to provide reliable insights into traffic conditions. The contracted tender of FVD provided the following data coverage on their provided map:
    - Segments with a maximum length of 50 m
    - Every road in the Netherlands + 10 km across the border
    - Every month a new basemap is provided that closely reflects real-world conditions. A lot of map-matching work between data is performed once to integrate a new map into the NDW system. This is useful to monitor metrics over a longer period.
- Every minute, NDW receives a file with the speeds (and travel times) of all FVD segments for which a speed was calculated in the past minute. The contracted tender of FVD provided the following data completeness on their provided map:
    - The vehicle speed value of the segments (first line of the data packet contains the delivery timestamp in UTC and the basemap which the segment division is based on)
    - SegmentID: A reference to the relevant segment
    - TraveltimeMS: The calculated travel time over the entire segment in milliseconds
    - Update frequency: Every minute
    - SpeedKph: The calculated speed over the segment in km/h

14 https://docs.ndw.nu/en/

- LOS-Reference: The level of service of the segment, a percentage from 0 to 100, where 100 indicates that the calculated speed is at least as high as the optimal speed of the segment from the basemap
- Only get data when a value of a probe vehicle is measured withing the last 30 minutes; only get the last 10 probe vehicle data
- Coverage bitstring, see Appendix C: Award criteria specifics:
  - The number of vehicles (1 to 10) which the speed value is based on.
  - In the past 30 minutes (using binary code to describe 5-minute intervals), in which period was the vehicle data measured? This was required in the second tender to indicate the data's age (simply, how old the data is). A bitstring indicating how many vehicles were measured 0-5 minutes ago, 5-10 minutes ago, up to 25-30 minutes ago. The total number is a maximum of 10.

- Accuracy of traffic speed and travel time: The RMSE and MAPE are calculated to evaluate the precision of the FVD data. These metrics measure the deviation between the FVD-provided speeds and travel times versus those captured by a high-accuracy reference system, such as the AVG (Actuele VerkeersGegevens - Current Traffic Data).
- Timeliness: Timeliness is measured in latency (as low as possible). The data stream's ability to reflect current traffic conditions in near real-time is crucial for its utility in dynamic traffic management. The timeliness score in Table 5, Section 5.1.2, evaluates the delay between the actual traffic event and its detection in the FVD stream.
- Quality of data delivery: This includes the reliability of data transmission and the extent to which the data stream meets the quality benchmarks set forth by the NDW. This is particularly important in scenarios where quick response times are necessary to manage traffic flow and prevent congestion.

The above-mentioned parameters were selected based on familiarity from a traffic engineering perspective, alignment with legacy systems for traffic management purposes and/or by experience gained through previous procurement and utilisation of 3rd party data in combination with verification requirements.

Applying the FVD algorithm to the above procured FVD stream, the aggregated NDW data are produced from segment level to route/trajectory level. The aggregated data can be found on open data portal in, for example, xml, json, geojson and csv format. NDW makes these data sets public and free of use since this is not the procured FVD data stream anymore and it has been aggregated by NDW.

The ROMO data provides an innovative and experimental methodology of NRAs procuring third-party data, particularly in the context of asset management and infrastructure monitoring.

From the workshop, we dive into the research questions defined in Table 1, the outcomes and lessons learned show that ROMO1 focuses on the use case description and functional requirements instead of data sets itself. Therefore, sharing of data stream is not applicable here. Instead, Mercedes-Benz has validated the data sets using a conceptual methodology proposed in the ROMO paper on Appendix D, which outlines their approach to assessing the quality and applicability of the probe vehicle data to meet specific criteria relevant to road condition monitoring and maintenance planning. The validation process involves:

- Vehicle sensor data Mercedes-Benz vehicle: Utilise vehicle sensor data to create a pseudo road profile.
- Data noise filtering: Apply frequency analysis to separate road surface characteristics from vehicle noise.
- Data accuracy and reliability:
  - Test with a fleet of vehicles across the Netherlands and compare with traditional methods.
  - Similar to the FVD stream, the accuracy of the ROMO data is assessed using

statistical measures that compare the data against a reference standard to examine its reliably for decision-making by the road operators.

By employing a conceptual validation approach, this section further explores the FVD and ROMO data streams' attributes and applicability can be investigated. The insights gained from the exploration of the FVD and ROMO data streams validation provide further confirmation of the challenges and opportunities associated with third-party data usage by NRAs. By conceptually evaluating these data streams against stringent criteria, this section confirms their current utility.

6 Main lessons learned

Building on workshop findings from Section 4 and 5, Section 6 presents the main lessons learned by addressing the set of research questions defined in Section 3.2. Procuring third-party data differs significantly from purchasing traditional infrastructure like asphalt. When procuring data, key consideration must be taken into account to align the effectiveness of the data for its intended (multitude) use (which might even include further distribution to other parties), whereas traditional infrastructure usually has a predefined clear objective. What also separates procurement of data from traditional infrastructure, is that data usually serves or facilitates working processes that, because of the availability of data, will change. Therefore, data procurement usually goes hand in hand with organisational / procedural changes.

These main lessons learned directly address the objectives of D4.1 by:

*Objective **1** (What?)*: Identifying key considerations in NRAs' practices and experiences across all phases of third-party data usage. The overview of the consolidated key considerations applied by the selected NRAs during the five key phases, answer the “What?” (presented in Italics in this section).

*Objective **2** (How?)*: Examining how NRAs approach third-party data procurement, integration and management. The detailed outcomes in the GRM tables highlight specific goal setting process, data integration techniques, data management methods, procurement models and strategies used by the selected NRAs to tackle operational, technical, and regulatory challenges (see Appendix E). These insights offer practical, hands-on advice for effectively tackling these issues.

*Objective **3** (So what?)*: Synthesising actionable lessons learned from the workshops and desktop research. The lessons learned discussed below focus on practical recommendations. NRAs can reference the knowledge to embark on implementing third-party data solutions more efficiently in traffic management, asset management, and road safety.

Pre-procurement phase

- Pre-studies such as research, trials, initiatives, proof-of-concepts, and pilots are essential in the pre-procurement phase, to find the criteria that define the goals of third-party data procurement. These pre-studies are typically smaller in scale, wireframe-based, with limited scope and (descriptive) functionality. Their purpose is to provide insights for potential procurements, which may occur years later.
- Market Consultation (MC) plays a vital role in helping NRAs gain valuable insights into the capabilities and offerings of third-party data providers. By engaging in MC, NRAs are better equipped to understand the latest technologies, market trends, and available competencies, which enables them to make more informed and strategic decisions during the procurement phase. This process ensures that NRAs can tailor their data needs effectively, select the most suitable providers, and optimize their procurement strategies to improve long-term outcomes.
- Gathered from the current practice of NL NRA third-party data procurement during the workshops, the criteria to determine third-party data procurement goals are based on validating the effectiveness of data for these specific goals, considering both technological innovation and budget feasibility. For innovative solutions, these criteria shall embrace continuous improvement to enhance traffic and asset management.
- Stakeholder engagement is critical for effective traffic data management, ensuring collaboration and alignment between various parties involved in the pre-procurement, procurement, and implementation of data. Strong communication and cooperation between stakeholders help NRAs make better decisions and deliver high-quality services utilising third-party data.

Procurement phase

- Taken from the workshops, third-party data procurement offers a strategic approach to overcome limitations in traditional traffic monitoring systems. By leveraging external data sources, NRAs can enhance data coverage, improve data reliability, and reduce data collection costs. This approach supports more effective traffic management, road safety, and infrastructure planning, all while addressing budget constraints and the need for scalable, innovative solutions.
- Setting **minimum requirements** for third-party data during the procurement phase is crucial to ensure data quality and integrity. These requirements help guarantee that the data aligns with the organization's standards and fits seamlessly into existing systems. Moreover, establishing these minimum thresholds prevents low-quality datasets from being accepted, even if they come at a low price. This prevents cost-driven procurement from undermining data reliability and ensures that quality takes priority over cost.
- Establishing **award criteria** during the procurement phase is crucial for ensuring both high data quality and cost-effectiveness. These criteria provide a structured approach to evaluate bids, where the highest score wins, ensuring that bids are assessed based on a combination of quality and cost factors. This method ensures a balanced consideration of both data accuracy and the financial aspects of the bids.
- Procurement award models offer different methods for balancing price and quality when selecting third-party data. The current common model is BPQR, which considers both cost and quality, ensuring a fair evaluation process. It is more comprehensive compared to models like Lowest Price, which focus solely on cost and risk sacrificing quality. BPQR is widely regarded as a more balanced approach, ensuring high-quality and maintaining competitive pricing.
- Data licensing in public procurement is vital for NRAs, as it determines who can access, share, use and monetize the data. NRAs often seek agreements that prevent third-party data providers from commercially exploiting the data, ensuring it is primarily used for public interest. However, negotiations vary, and some NRAs may impose strict limits, such as prohibiting redistribution. At a minimum, NRAs require control over data usage and aim to balance public benefit with data security.

Implementation phase

- Data attributes align with the requirements and criteria during the procurement phase and are critical in the implementation phase. Key data attributes like accuracy, coverage, and update frequency ensure seamless integration into existing systems and fulfilment of operational needs. Defining these attributes early guarantees reliable data delivery that supports real-time traffic management and decision-making.
- It is key to actually **understand the data** that is procured. This is especially the case when data has not been used widely before or if the data service provider is launching a new service and is also not experienced with what the data actually represents.

As an example, in the ROMO project, Mercedes provided data that indicated potholes in the road. It took quite some iterations between Mercedes and NDW before the right interpretation of the data was established. At first, the data suggested that Dutch roads were flooded with potential potholes. After analysis, it also turned out that different car models produced different data profiles for potholes. Therefore, a lot of filtering, etc. had to be done. This phase should not be underestimated. Obviously, some data is more prone to discussion than other, e.g. there is probably less discussion about “outside temperature”.

- Effective data integration is essential for combining diverse data sources to meet complex project objectives. It ensures accuracy, consistency, and alignment with system requirements, enabling seamless data integration with existing data. Plans for how to integrate third-party data with existing ones (e.g., systems, APIs, databases) must be considered in previous phases before implementation phase to have strong project outcomes.

- System operation and data extraction of procured data require ensuring compatibility with existing platforms and proper adaptation for smooth integration. Aligning with internal systems and needs, which involves significant adjustments, procured data is pulled or pushed. The data extraction is usually tailored for maintaining performance, accuracy, and compliance with data protection regulations. Early planning during procurement reduces the effort needed for adaptation and ensures data utility during implementation.
- Data access control for procured data and its derivatives must balance open access for public use and controlled distribution for private entities. Procured by governments, governmental organizations should be given free access to improve traffic management and public safety, while private companies may access it via paid licenses. This model promotes data usage of NRAs while ensuring that the data's value is preserved and incentivizes innovation in traffic solutions.

Maintenance and monitoring phase

- Maintenance and monitoring must focus on maintaining high data quality throughout the operation. This includes regularly ensuring that metrics such as accuracy, coverage, and timeliness stay aligned with the KPIs outlined in the award criteria during the procurement phase. Regular monitoring ensures that the data continues to meet operational standards, enabling effective real-time traffic management and long-term system effectiveness.
- Effective communication and regular collaboration ensure that stakeholders' evolving needs and issues are addressed and that the system remains responsive during operation, maintenance and monitoring. Establishing clear communication protocols facilitates smoother navigation of complex data pools and ensures that the system is adaptable to changing requirements.
- Monitoring protocols ensure that problems observed such as system disruptions or data inaccuracies, are identified and addressed promptly, maintaining overall system performance.

Overarching phase: SLA, scalability, challenges and risk

Third-party data procurement

- SLA monitoring focuses on ensuring that key service metrics—such as data quality, uptime or downtime, latency, and response times—are continuously evaluated and meet the standards. Regular testing of the procured third-party data and reporting up-to date allow for early detection of issues, maintaining service reliability and operational efficiency.
- Scalability is one of the key considerations when expanding infrastructure coverage for data-driven road management. Traditional detection methods can be prohibitively expensive or unfeasible, especially for large-scale or real-time monitoring of road management. Scalable in-house data systems need to process large volumes of data efficiently and extend coverage beyond main roads to secondary or rural areas.
- Another key consideration is **continuity**. During the procurement phase NRAs have to think about requirements that mitigate the risk that the third-party service provider will not stop the flow of data somewhere during the maintenance and monitoring phase. Since it might be a new business, it may turn out not to be profitable for the contracted partner to continue the services. Therefore, the NRA has to think about a fallback or

incentive to not let this happen (or at least to diminish the chance of this happening), e.g. financial consequences if third-party service providers terminate the service before the end of the contracted period.

- Challenges in third-party data procurement vary significantly, ranging from licensing restrictions to organizational complexities. Procurement challenges often stem from the upscaling of data management, internal resistance, high cost on in-house systems development, data collection/integration difficulty across systems and systems update due to change of third-party data vendors.
- In third-party data procurement, both anticipated and unforeseen risks arise, such as cybersecurity, regulatory concerns, extra cost for system update etc. These risks need mitigation strategies to ensure the continued integrity and usefulness of the data while avoiding system disruptions.

Comparing to third-party data procurements, traditional procurements such as road infrastructure and traditional detections procurements have other risks.

- Procuring road infrastructure is expensive, time-consuming, and susceptible to regional changes, requiring careful planning and risk mitigation strategies to prevent delays and cost overruns.
- Another major risk is cost overruns, which are frequent in large-scale infrastructure projects due to various factors like poor project planning, design changes, and unforeseen technical difficulties. For example, materials cost fluctuations and scope changes can increase expenses significantly. These projects also face delays, with technical challenges such as coordination between different stakeholders and labour issues contributing to missed deadlines.
- Additionally, infrastructure projects are often at risk of becoming outdated by the time they are completed, particularly in rapidly changing regions. The long timelines for road construction can lead to traditional detections designs that no longer meet current needs by the time the project is finished. This is compounded by the difficulty of coordinating multiple contractors and stakeholders, further complicating efforts to deliver on time and within budget.

D4.1 helped to realise that the lessons learned are not only preserved but are also mapped into actionable advice for NRAs across Europe. The advice corresponding to use cases for NRAs to consult when embarking on third-party data procurement initiatives is a crucial step in advancing the PRESORT project's goal of empowering NRAs to efficiently acquire and utilize third-party data.

Next steps

The most important lessons learned is that procuring third party data needs a different approach from traditional infrastructure procurement. NRAs should, when considering procuring third party data, focus on the purpose of the data once retrieved, the goal it is intended for and organisational and procedural impact it will generate. Two items were identified that need further attention (in the next Work Package 5) because of the impact it bears on all phases: which licence does the NRA need to allow use of the third-party data as foreseen (e.g. use by other governmental institutions) and which procurement model is best suited to actually acquire the data that is needed?

References

[1] Floating Car Data Report Document Information, FCD Workshop, Frankfurt 17th September 2019. Retrieved Oct 3, 2024, from https://www.its-platform.eu/wp-content/uploads/ITS-Platform/CorridorDocuments/UrsaMajor/UMneo_FCD_WS_Report.pdf

[2] ITS in the Netherlands Progress Report 2010 – 2014. Retrieved Oct 3, 2024, from https://transport.ec.europa.eu/system/files/2016-09/2014_nl_its_report_2014_en.pdf

[3] Viti, Francesco & Hoogendoorn, Serge & Immers, Ben & Tampère, Chris & Lanser, Sascha. (2008). National Data Warehouse: How the Netherlands Is Creating a Reliable, Widespread, Accessible Data Bank for Traffic Information, Monitoring, and Road Network Control. Transportation Research Record. 2049. 10.3141/2049-21.

[4] Request for Quotation Floating Car Data Netherlands In the context of the Framework Agreement for the Supply of Traffic Data from Non-Road-based Systems, Retrieved Oct 3, 2024, from https://www.tenderned.nl/aankondigingen/overzicht/264689

[5] Commission Delegated Regulation (EU) No 886/2013 of 15 May 2013 supplementing Directive 2010/40/EU of the European Parliament and of the Council with regard to data and procedures for the provision, where possible, of road safety-related minimum universal traffic information free of charge to users Text with EEA relevance,), Retrieved Oct 3, 2024, from https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=celex%3A32013R0886

[6] Public European Tendering Procedure Descriptive Document – ROMO NDW Agreement for Road Monitor (ROMO), Retrieved Oct 3, 2024, from https://www.tenderned.nl/aankondigingen/overzicht/318103

[7] European Road Safety Observatory__Statistics and analysis archive__Esafety__ECall, Retrieved Oct 3, 2024, from https://road-safety.transport.ec.europa.eu/european-road-safety-observatory/statistics-and-analysis-archive/esafety/ecall_en

[8] Emergency-Call-e-Call-Initiative-in-Europe, Retrieved Oct 3, 2024, from https://sesei.eu/wp-content/uploads/2019/05/Emergency-Call-e-Call-Initiative-in-Europe.pdf

[9] EU Road Safety Policy Framework 2021-2030 - Next steps towards "Vision Zero". Retrieved Oct 3, 2024, from https://op.europa.eu/en/publication-detail/-/publication/d7ee4b58-4bc5-11ea-8aa5-01aa75ed71a1

[10] Directive 2014/24/EU on public procurement and most economically advantageous tender (MEAT) criteria, Retrieved Oct 3, 2024, from https://www.europarl.europa.eu/doceo/document/E-9-2020-005638_EN.html

[11] Consolidated text: Directive 2014/24/EU of the European Parliament and of the Council of 26 February 2014 on public procurement and repealing Directive 2004/18/EC (Text with EEA relevance)Text with EEA relevance, Retrieved Oct 3, 2024, from https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A02014L0024-20220101

[12] Directive 2014/25 - Procurement by entities operating in the water, energy, transport and postal services sectors, Retrieved Oct 3, 2024, from https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A02014L0025-20220101

[13] Statement of Work (SOW FVD 1.4) Het leveren van actuele verkeersgegevens op basis van Floating Car Data aan NDW, Retrieved Oct 3, 2024, from https://www.tenderned.nl/aankondigingen/overzicht/264689

[14] Producten Floating Car Data (FCD)—NDW Docs. (2024). Retrieved Oct 3, 2024, from https://docs.ndw.nu/en/producten/FVD/

Appendix A: Report on workshops with the Netherlands NRA

The Appendix A presents the synthesised workshop logs with the Netherlands NRA, focusing on two third-party data use cases: FVD (all three tenders, especially on the third one) and ROMO (covering ROMO1 and scalability in ongoing ROMO2).

The main objectives of these two workshops include:

1) Examine the pre-procurement, procurement, implementation, and maintenance and monitoring of third-party data in traffic management and asset management.

2) Share experience and insights regarding overarching phase on the data quality, SLA, scalability, barriers/challenges and risk in current practices.

3) From 1) and 2), recognise the findings and the underlying reasons, and present the outcomes and lessons learned to NRAs as outlined in GRM of Section 6.

During preparation, both workshops employed the GRT approach to identify the phases and key considerations for each phase. This process helped formulate research questions for the workshops and their respective parallel sessions (effectively two separate workshops). Appendix A addresses these questions directly. The following workshop logs provide a detailed account of the discussions (each row one discussion), though the conversations may not always appear in chronological order.

Note: *Disclaimer* is stated in Section 4.3 Workshop log

Appendix A1: FVD workshop

Table 6 FVD workshop log synthesised from the workshop transcript

Agenda Timeline 0	Nuno Rodrigues (NR-MAPtm)	Shubham Soni (SS-RHDHV)	Tomi Laine (TL-Traficon)	Jilt Reitsma (JR-NDW)	Marlous Hovestad (MH-NDW)	Key insights	Takeaways
Rationale for FVD procurement 1	Asking research questions (defined using Guidance Reference Table approach) one by one				Why procure FVD	Some roadside equipment but not enough for whole coverage	-Loop detection expensive, values fluctuate strongly in time -Not flexible and scalable -FVD can increase the coverage on travel time trajectories -FVD can minimise the data collection cost
2			Difference to Finland NRA			Procured third-party data was unusual to made public	Off-the-shelf data was procured to target this.
FVD goals Criteria 3	Raise the research question			Travel time	pre-study	A pre-study (the data quality method) to run some tests (not real-time data)	The criteria were developed based on research studies and pre-study (trials).
			Market consultation (MC)	Experts/ competencies availability Understanding of third-party data	Learns more about data algorithms	Via MC, more understanding of the technology competency, availability, data algorithms are gained.	MC rationale - Technology competency, availability, data algorithm - NRAs Making better decisions on third-party data, e.g. NDW takes on the updating maps from third-party data providers, performs map-matching and building an in-house system with its own

Conférence Européenne des Directeurs des Routes
Conference of European Directors of Roads

4							team so that changing to a new third-party data provider would be possible. - With deep understanding from MC, NRAs are in better position to adjust and take out what data they need - Knowing the market and potential companies that can provide and bid
FVD procurement				Zuid-Holland Pilot success; FVD Three tenders		Zuid-Holland, a pioneering pilot project was initiated, which incorporated ANPR data into the traffic management system to enhance real-time traffic monitoring and optimize the accuracy of travel time calculations by combining ANPR data with other traffic information sources. The pilot was a success and laid the groundwork for the broader application of data-driven traffic management across the Netherlands, contributing valuable insights to subsequent procurement processes. Three procurement	**First** tender: - Ask out: 2016 - What: pilot for the Province of South-Holland, with an option for the Netherlands - In production: halfway 2017 - Supplier: Be-Mobile **Second** tender: - Ask out: 2020 - What: the Netherlands - In production: April 2021 - Supplier: Be-Mobile **Third** tender: - Ask out: at this moment - What: the Netherlands whole network - In production: N.A. - Supplier:

5						tenders of FVD to enhance its traffic management and monitoring capabilities. The first two tenders, initiated to source real-time traffic data from connected vehicles, aimed to improve road safety, optimize traffic flow, and provide more accurate travel time predictions using FVD.	
6					Off-the-shelf data	NDW needs to be transparent and explain how the NDW data were generated. It is more convenient to procure off-the-shelf third-party data, so that integration and aggregation to in-house system using NDW's own algorithms are explainable.	Off-the-shelf data is the lowest cost. NDW in-house system is the good foundation to support this kind of procurement.
					Data quality test	NDW asks for real-time data feed of two weeks to test and avoid other sources of added data that third-party data providers retrieve or aggregated via other channels. NDW determines the ground truth based on its road detector data, camera data and other available in-house data, so that all third-party providers' data can be	**Tender methods** - NDW Ground truth - Minimum requirements for the tests, such as strictly real-time data etc - Using speed heatmap to see how much FVD data overlap with NDW ground truth **Availability** Scope - Time: update frequency

7						compared to the ground truth (benchmark, not acclaimed as the real ground truth).	- Space: smallest available segments and their information, all road networks Content - Data types: speed - data format: OpenLR, GeoJason **Timeliness** - lowest data latency possible - lowest data age possible **Quality score**: - QKZ-method: Determine the congestion matches between FVD and NDW ground truth - RMSE (root mean square) and MAPE (mean absolute percentage error) are both essential to determine the data quality
8	Award Criteria and other key considerations to evaluate the data sets				Procurement documents	The procurement process matures throughout the years of third-party data tendering history.	• Lessons learned process from first tender (earlier data qualification criteria) • Procurement documents specify - Tender model and methods - Value calculation for every part
Implementation phase					Awarded bidder data attributes	NDW set the requirements and data quality test of the tender.	What data attributes does the awarded bidder offer? Delivery:

9							- Every minute - Only data when a GPS position value is measured - Speed - Coverage Map: - Segments with a maximum length of 50m - Every road in the Netherlands+10km across the border - Every month a new map, a special team in NDW to perform map-matching integration. With successful map matching, the average speeds is calculated.
					Off-the-shelf data into NDW aggregated products	NDW bought FVD data aiming for the real time travel times on predefined trajectories (around 100,000 now). NDW's real-time system can flush out the travel times on all the predefined directories every minute. It has two displays; by clicking on one trajectory, you can get historical travel times if the real-time one doesn't exist. A new feature, which shows real-time	NDW can make their own product: - Realtime travel times for predefined trajectories - Historical travel times for flexible trajectories - Realtime congestion information - Safety indicators, e.g. estimated V85 - Aggregated speeds per month / year per segment - Speed heatmap for road segments

10						congestion information. NDW made an algorithm to determine where the start and the end of the traffic jam is, and this data feed is used in the traffic management centre. Estimated V85, 85% percentile of all the speed can be used for safety indicator. If a lot of vehicles are speeding based on this indicator, that's an indicator that used for safety on this trajectory.	- For governments, governmental organizations, this access is **free**. Other companies, such as consultancy company can use the tooling/data through a **license** with extra conditions
11	Implementation process on integrating, managing and monitoring				Pulling instead of pushing data Two in-house systems		When - Aggregating the procured FVD, data pulling from the procured data - Make sure targeted data are retrieved Two in-house systems were made: - Aggregating data - Hosting trajectories: Trajectenmaker
	Open Data Portal				NDW products using procured FVD are made, then add such as cause of delay etc. by the TMC team.		- Open Data Portal is public and free to use: Derive from NDW products which derived from procured FVD; congestion data from other data sources, are merged into TMC system - NDW products for governmental

12							organisation or companies (licencing). TMC have access of the NDW product “Viewer” to see the real-time traffic information on map (no integration to TMC system)
Maintenance and monitoring phase 13				Maintenance is performed in-house	Issues	-Pulling data in rush hour	Maintenance and monitoring issues - Pulling data of a requested timestamp during rush hour may not be processed in time - Solution: pulling with “second zero”, “second one”... to get the ”first zero” ”first one”, pull in 10 sec, sent out within 50 sec. - Needs to handle in care, good to have a in-house system to maintain
SLA, scalability, challenges and risk 14		scalability			Data size	NL: 10 million segments	- Data sets are kept as small as possible - Only send out data on the segments where there is data - Some providers take out data headers to minimise the size of data, might cause potential risks to data interpretation, integration and aggregation

				Challenges		Not a lot of companies want to bid on providing off-the-shelf data. They want to make their own product based on their data to benefit from licencing in the future.	Challenges: - Licensing issues for open access to aggregated data. - Segment data lacks OD (Origin-Destination) information, prompting a separate tender process. With OD data procurement – took a long time to start procurement due to slow organizational process. - Tender issued during the COVID pandemic, almost no congestion --> is necessary for the data quality tests. - Systems need to handle massive data volumes. Due to the immense data size (a day of minute data (1440 datasets) could be 10Gigabyte,), a scalable system is required to process data in real-time. - Risk and mitigation: ○ Requirement of a stable fleet to provide FVD

15							○ Exigent situations such as bankruptcy can happen to both third-party data providers and loop detector data providers ○ Policy and regulation
16					In-house system	NDW builds in-house data systems to perform data processing, aggregating, product developing.	In-house system - Costly but to make sure NRAs are in control, are independent from one provider, and are able to develop its own products for the stakeholders - In-house system needs adjustment when third-party data provider is changed - Sometimes provider's data algorithm, which causes SLA changes, so that difference outcomes are explainable to NDW's partners.
	Risks such as security, privacy etc				Special data Special API GDPR	Third-party data provider only provides the data, but not allowed to	FVD is better than ANPR and Bluetooth connection. Procurement documents

17					compliant	access NDW in-house system	specified directives and rules a provider must follow, by checking such boxes, potential risks are prevented on - Special data, API for security reason - GDPR for data privacy protection
18				Off-the-shelf FVD	in-house system development	From maintenance/monitoring, Anticipating in-house system updates	-Blocking the NDW development team for the last quarter to perform potential in-house system update -largest risk is change of supplier (Time and cost risk)
19					Circumstances COVID	Insufficient data	-contingency plan to mitigate in extenuating circumstances
20	SLA monitoring				Quality test monthly according to the quality standards		SLA monitoring: - Use the same data quality methods as in tender methods - Test frequency, monthly - On spreading trajectories, not on the whole network since there is no real ground truth of the whole network - NRA internal team: Data quality on the reference system, a fixed subset of loop detector, exchangeable to another subset when the data quality is not ok.

21				Quarterly reporting on SLA			Quarterly reporting on SLA - To catch on issues or potential escalation - To meet the threshold of uptime and downtime, e.g., Downtime shall be less than 8 hours per year
22			Tendering frequency and length		Different procurement to make the NL attractive for other projects	Contract period 2 years, with 2 times a 1-year extension procurement frequency	- Longer contract is not desirable to make market attractive
23	Organization issues				NDW is responsible for all its partners.	Since the data is provided by NDW to other road operators or govt., for third-party data providers there is no demand from that side.	NDW is procuring in a centralised way on behalf of its governmental organizations: -Third-party data provider values this model as it is time saving -Data supplier disappearing from market once the large tender is settled.
24	Benefit and cost	Benefits Coverage and costs, Whole country within one contract			Procuring the whole road network	Round number of costs is difficult to estimate since it is unrealistic to cover all roads with traditional equipment. Cost for building and maintaining the in-house system is also difficult to estimate. Adjustment of system costed around 500k (approx. using rule of thumb) euros in 2017.	Benefits: - Saving cost coverage using traditional road equipment implementation and maintenance - New data products such as V85 to provide estimated speeding information to highlight the safety issues.

Appendix A2: ROMO workshop

Table 7 ROMO workshop log synthesised from the workshop transcript

Agenda Timeline 0	**Giovanni Huisken (GH-MAPtm) Xiaoyun Zhang (XZ-MAPtm)**	**Boris Kock (BK-PNH)**	**Scott Stephenson (SS-AECOM)**	**Ernst Scheerder (ES-NDW)**	**Erik Vrijens (EV-IenW)**	**Key insights**	**Takeaways**
Rationale for ROMO procurement 1	Asking research questions (defined using Guidance Reference Table approach) one by one				Using probe vehicle data for asset management	Using free probe vehicle in-vehicle data for safety purposes, procuring data for road management purposes. 4-5 small-scale POCs (Noord-Brabant etc.): Benefits in probe vehicle in-vehicle data: the exchange of data provides great opportunities in terms of traffic management, traffic information, and incident management, with positive effects on traffic flow, safety, and car traffic emissions	Rationale - More budget constraints to develop physical infrastructure - Increase RM coverage - Look for innovative /alternative solutions Therefore, 10 Uses cases were identified in three categories: Asset management Winter management Road safety
ROMO goal Criteria					The goal of the procurement	To see the validity of using probe vehicle data for asset management.	- Open to dive into new technology helping us - Find the budget to do so - Non quantifiable Criteria to validate whether the procured data met the goal.

2							- Goals setting is flexible with innovative solution
ROMO1 Procurement 3					One provider, Mercede-Benz secured all three lots in the tender.	A strategic series of bids is offered by the winner, to show their active involvement in this field.	- Strategic bids across three categories - Data aggregation is typically handled by a separate division among car manufacturers, and many rely on separate companies for this process. However, Mercedes-Benz manages its own vehicle data aggregation in-house. - Giving the "useless" data, such as missing lanes information detected by Lane Keeping System, a second life via ROMO
4					Price-quality procurement model	A very low price yet with a high data quality	- Strategic on the price to position on the markets - Advise a non-traditional procurement model like ROMO1 - Procure data from multiple car manufactures
Implementation					Drivers need to opt in for their data sharing to car manufactures	While safety related data are obliged to be opted in, ROMO data implementation requires the driver to opt in for sharing these data with Mercede-Benz, to contribute to TM and	- Consent of data sharing from driver to Mercede-Benz - GDPR, strict personal data protection - Third-party data provider performs data anonymisation GDPR Article 29

5						RM.	defines personal data, anonymous data, and pseudonymised anonymous data
6					Communication between stakeholders	Car manufacture engineers and road operators from NL NRA were building an improvement communication schema	- A collaboration scheme to regularly reflects the needs of stakeholders, the issues at hand and possible solutions
7					How to navigate a large data pool	Methods to "swim" in a large data pool is crucial.	- Finding the safety hotspots based the data - Finding the safety critical events
8					Difference operators, such as city, province, regio etc, have different levels of data competencies		- Solutions provided for the NRAs needs to be intuitive and streamlined.
9					Dashboard	Displayed instead of further integrate. Dashboard is the second screen, once confident, integrate to NDW existing in-house system.	- Visualised for certain event - Dashboard shall show the timestamp of certain event - After confidence check of what the dashboard is showing, integrate to NDW product- the viewer - NH has its own API connection with the Dashboard
Maintenance and monitoring					Sub-optimal event	ROMO1 was not set to be technical, but then we miss specific	- Goal was not a technical project, but a descriptive and data

10						attributes. More specific data attribute was in need later on.	presented in dashboard - Road segment information: slippery segments and the temperature of the segments (which was not considered during the procurement)
11					Quarterly report, payment		- Quarterly report to monitor, check issues and confirm payment
SLA, scalability, challenges and risk 12					SLA metrics		- Uptime, downtime of API - Response time of the dashboard when a request is made - Data delivery latency is set to minimum requirements
		Challenges, action list			Underestimate the degree of organization collaboration complexity		- Car manufacture engineer and road operators' communication and collaboration - Trade-off between Data refinement and pushing to Dashboard - Road operator involved is not on ROMO payroll. They contribute to analysis on volunteering base. - Road operator (municipal, regional) wants a straightforward action list to tackle the events from 1-10 - Strict budget on road infrastructure

13							maintenance, which cause internal resistance to perform innovative project like ROMO
14					ROMO2 rationale ROMO2 budget is approved	- ROMO1 turns out to be a POC development, ROMO2 is needed to go into specific data sets - Prevent vendor locking	Before ROMO2, A market consultation was performed to prepare for it.
15					ROMO2 concept	Similar use cases Similar technologies Similar dashboard approach, but multiple data sources, which integrate, aggregate and complement each other	
16				Car manufactures roadmap NDW data scientist		Car manufactures won't change their roadmap, they have data and that's what they can provide. NDW data scientist has difficulty looking at these new, unstandardised data	Asking from car manufactures specific data from a certain sensor, the communication can be tricky - Internal discussion of rationale for the car manufactures to do this - Partnership but not contractor relation between NRA and car manufactures
					Prevent vendor locking	To prevent vendor locking, we avoid data providers forming alliance but encourage multiple data sources.	- Multiple tenders - Each use case →a procurement - Each procurement→ multiple data sources

17							
18					Road management standardisation	TNO NL is working on the research to promote the standardisation of road management	Road and asset management standardisation is quite low regarding the innovative data exploitation.

19					Practical and productive procurement		- Consult procurement lawyers on how to procure in a productive way.
							- Procurement framework
							    - Variable renumeration
							    - Renumeration of a data source is dependent on the data volume and data quality

20			Data quality integrity		Investigating Variable renumeration	Volume quality matrix	An innovative procurement framework is needed - Different car manufactures have different detector levels which produces different data quality
21					GDPR Cyber security		- Car manufactures have strict monitoring of GDPR and security - GDPR and API cyber security stated in the procurement docs - Road operator standards used as added-value optional in the procurement docs

Appendix B: Report on workshop with the Ireland NRA

Appendix B summarises the synthesised workshop logs with the Ireland NRA on third-party data and eCall data pre-procurement, (potential) procurement and implementation. The workshop objectives include 1) explore the current data pre-procurement (potential) procurement of using third-party data and eCall data in Ireland; 2) exchange and share experience and insights to enhance the awareness of third-party data usage 3) recognise the barriers/challenges, and risks of third-party data of and eCall data to demonstrate the lessons learned to other NRAs. Note: *Disclaimer* is stated in Section 4.3 Workshop log.

Table 8 Third-party data and eCall data workshop log synthesised from the workshop transcript

Agenda Timeline 0	**Giovanni Huisken (GH-MAPtm) Xiaoyun Zhang (XZ-MAPtm)**	**David Laoide-Kemp (DL-TII)**	**Daniel Pentony (DP-TII)**	**Cormac Synnott (CS-TII)**	**Danny Woolard (DW-Vesos)**	**Andy Graham (AG-WhiteWillow)**	**Key insights**	**Takeaways**
Welcome 1	Host, Everyone						Introduction of PRESORT project	
Third-party data 2		Data collection					- Traditionally IE NRA collected own data essentially through NPR cameras or loop detectors - No specific policy to explore third-party data sources yet - IE NRA has been approached	- Third-party data is getting attention recent years - Third-party data is cheaper in many business cases comparing to traditionally collected data

							by third-party data providers	

Current status Pre-procurement			Trial				Trials (WAZE and Valerann)[15][16]in pre-procurement phase are done to examine:
							- Data cost
							- Data accuracy
								- Trials/Pilots need to be in place to study data **cost** and **accuracy**
								
								- Two-way exchange:
								  - Consider data from WAZE crowdsourced app (stopped vehicle data are way off carriage way, inaccurate data can't be used by TMC)
								  - Investigating the benefits of data fusion between NRA in-house data and third-party data.
								    ○ six months subscriptions
								    ○ Fused data aim to show a real time overview of the

15 https://irfnet.ch/2024/04/09/valerann-to-showcase-the-benefits-of-data-fusion-for-road-safety-and-incident-management-enhancement-on-irish-motorways/

16 https://www.valerann.com/news/valerann-to-showcase-the-benefits-of-data-fusion-for-road-safety-and-incident-management-enhancement-on-irish-motorways-2

3								traffic situation, road safety, Inc. events and incident alerts - Two categories of roads to implement trials and study benefits o Average amount of ITS with limited cameras, high traffic volume o no cameras and low traffic volume
4			Criteria to define the general third-party on stopped vehicle utilisation					- Improve the detection and classification of road incidents - Reduce our response time to get our instant response units out - To reduce the time to incident recovery - To reduce time informing the public of the situation by messages and VMSs
(potential) Procurement		Assessment of trials first						- Trials assessment of third-party data is crucial, sufficient

5								quality is a must - The future steps for third-party data procurement are not in place yet
6				Choosing one over the other based on trial assessment?				- Trials assessment: choose one over the other? - Or need both data sources
7		NTA procuring travel time						- NTA covers public transport - Developing an app where they want to compare public transport time with private vehicle travel time
Barriers/ challenges, risks 8	Data fusing experience in the trials						Data integration	- Offline. Not been integrated into in-house data bases. - Consultant from separate company in our control centre who has access to the cameras and has been looking at the data and seeing in real time instance as it come in. - Looking at developing inaccurate information filtering mechanism
9							How to procure How to evaluate data sets	
					European five-star			- NL NRA, especially

10					rating, data quality assurance guidelines and recommendations			NDW is vocal in data quality assurance
eCall 11						Presentation	- eCall button Mandated in 2018	- Airbags and manual activations - Increase activation in 2023, 1285 across up the Republic of Ireland airbag activations.
Studies 12							Studies has been carried out to show the potential	Procurement phases are not applicable as eCall data is owned by IE member states emergency services and don't require procurement.
13					Next generation eCall		- EU consortium kick-off project regarding next generation eCall - PSAPS (Public safety answering point) and the way that the data comes work.	- This European project initiates a whole upgrade of eCall over the next few years - Scalability of eCall into PSAPS
Challenges						Data filtering	log data - filter out all the insufficient quality data - Map match it to the Tii network. - Taking all	Data filtering is essential to filter out false alarm and erroneous data. Hypothesis for the false alarm can be validated, then remove the false alarm data.

14							the rural roads, city roads inside Dublin away. - Clusters on relevant network show	
15						Benefits	- Not being exploited in traffic management	- Possible time saving on equipped motorways, - Roads with much less ITS equipment far more time saving - Looking at eCall as the traffic management data. - How we can prove eCall data as a business case
16						Data anonymisation		- eCall data are in its nature personal data - Further usage needs to follow data anonymisation in GDPR - This could go unexpectedly smooth as eCall data is almost a real-time system of emergency
17	Technology updates 2G, 3G					-2G carry on -Fitting 4G and 5G device to new vehicle	-Ireland and many Member States, 2G and 3G are being turned off eventual -3G has already been turned off, largely in the UK	- Many things (burglar alarms, safety alarms for senior person, governmental approved smart meters in the UK, therefore 2G will carry on for the moment

Appendix C: Award criteria specifics

During the procurement of FVD, the quality of several road segments was evaluated. Travel times derived from FVD were compared with travel times from ANPR cameras. Key metrics such as average deviation (RMSE and MAPE), the accuracy of delay detection (QKZ), and timeliness were assessed. Additionally, completeness—how many minutes data was provided—was analysed. The quality of FVD largely depends on coverage size and distribution. Therefore, understanding the number of probes used in a measurement is crucial for gauging quality. NDW calculates and provides quality indicators such as **availability**, **coverage**, and **timeliness** [13].

Availability Quality Indicator

The availability of a segment refers to the percentage of the segment for which data is delivered. Be-Mobile provides data for a segment if one or more vehicles have delivered data within the last 30 minutes. Data isn't provided for every segment every minute. Availability shows what percentage of the segment delivered data. If availability is below a certain threshold, no travel time or availability will be displayed.

Coverage Quality Indicator

Coverage indicates how many vehicles the calculated FVD travel time is based on, using a coverage bitstring. This value does not consider the freshness or availability of data. For each minute, only the coverage of segments that provided data is used. Be-Mobile delivers coverage between 0 and 10, and the outcome is scaled to reflect the number of vehicles contributing to the measurement.

Timeliness Quality Indicator

Timeliness reflects how recent the FVD data used to calculate travel time is. It uses coverage per 5-minute intervals from the coverage bitstring. For each segment, timeliness is calculated based on the data provided in the past 30 minutes, with a score indicating how fresh the data is. A timeliness score of 80 means that the data used is, on average, 5 to 10 minutes old. Timeliness tends to be higher during the day when more recent data is available and lower at night, when older data is often used.

The DATEX-II field used for timeliness in FVD travel times is "supplierCalculatedDataQuality."

More detailed information can be found on NDW docs[17]. The about is used to develop the Award criteria specifics. Each criterion specifies are shown below.

G1 Bid Amount

The participant must submit a bid amount for the delivery of the requested traffic data. This bid amount is binding. NDW has a maximum budget available for the assignment, which is not made public. It is NDW's choice not to disclose the maximum budget prior to final award to avoid influencing the market and to promote competition. In principle, the bid amount offered by the participants should not exceed this maximum budget. After receiving the bids, NDW will first evaluate the bids that do not exceed the stated maximum budget and are otherwise deemed valid. These bids will be evaluated based on G2 and G3.

If NDW receives one or more valid bids that do not exceed the maximum budget, bids that do exceed the maximum budget will not be considered.

If all valid bids exceed the maximum budget and NDW manages to secure additional budget for the assignment, then regardless of the budget overrun, an evaluation of all valid bids will

[17] https://docs.ndw.nu/en/producten/fcd/#datakwaliteit

take place. In that case, the evaluation will follow the same methodology as if one or more valid bids had been submitted that did not exceed the maximum budget. In such a case, the schedule mentioned in paragraph 1.5 may need to be adjusted. We will inform you of this in a timely manner.

G1 Quality of traffic data

Quality of traffic data are determined by comparing the time of detecting a traffic phenomenon in the FVD with the time of detection in NDW data

An example data stream of the live data feed for the period of two weeks from June 10, 2024, 0:00 hours to June 23, 23:59 hours at a 1-minute level.

These data must be delivered according to the guidelines in the statement of work of FVD [13] document, which is a separate document with detailed methods to measure the quality of traffic data.

Participants must explain the quality of the traffic data and delivery based on the requested performance indicators for award criteria G2.1 to G2.6. For this, participants must also provide a real-time data feed, including the necessary files to interpret the data, in the same manner as it will be delivered during the operational phase if the participant is awarded the contract.

These data must be delivered according to the guidelines in Appendix B and accompanied by a description of the offered traffic data and any quality indicators. Additionally, the bidder must have an employee available to answer any questions from NDW for the correct interpretation of the data feed during the implementation of the real-time data feed.

G2.4 Quality score RMSE & MAPE

The RMSE is a quality measure that represents the average difference between the AVG and FVD speeds. Here, AVG (Actuele VerkeersGegevens - Current Traffic Data) serves as the reference system, which has been proven to adequately represent reality.

$$\text{RMSE} = \sqrt{\sum_{k=1}^{K} \frac{(v_k^w - v_k^r)^2}{K}}$$

Where:

k: Symbol for delivery period

K: Total number of delivery periods

V_k^w: Converted speed for time period k using a high-accuracy measurement system

V_k^r: Converted speed for time period k from the current FVD travel time

Score:

A common criticism of RMSE is that outliers during special events are not always well represented. To address this, three RMSE scores are determined:

RMSE-overall score on speeds

RMSE-congestion moments on speeds

RMSE non-congestion moments on speeds

Minimum requirements:

RMSE overall score on speeds: maximum 15 km/h

RMSE congestion moments on speeds: maximum 20 km/h

RMSE non-congestion moments on speeds: maximum 15 km/h

When assessing the quality score RMSE, the following elements are considered:

- The smaller the deviation of the delivered data compared to the NDW data, the lower the RMSE value, and the better it is rated in the RMSE score.
- The various components of the RMSE score (overall, congestion, non-congestion) are assessed separately; the average of these components forms the overall RMSE score.
- A score of 500 is assigned to the supplier with the best score. The other suppliers receive a score proportionally.
- A score greater than 0 is only possible if the participant's value meets the minimum requirements as described in [13].
- If there are nearly equivalent RMSE values from two or more participants (all of whom meet the minimum requirements), these participants will receive the same score. A margin of 20% relative to the RMSE value of the best participant in this set of participants with nearly equivalent RMSE values will be applied. The best-determined set of participants will thus receive the score "excellent," the second-best scoring set of participants will receive "very good," and so on.

When assessing the quality score MAPE, the following elements are considered:

- The smaller the deviation of the delivered data compared to the NDW data, the lower the MAPE value, and the better it is rated in the MAPE score.
- A score of 1000 is assigned to the supplier with the best score. The other suppliers receive a score proportionally
- A score greater than 0 is only possible if the participant's value meets the minimum requirements as described in [13].

The MAPE is a quality test applied to travel time data. For this, the reference system consists of so-called ANPR-trajectory travel time data. To determine travel times at the trajectory level from the supplied FVD, a "Trajectories method" is used. These two travel times will then be compared using a MAPE score. We only look at the period between 06:00 and 21:00. For each trajectory per day, we check if the MAPE score is below 20% (maximum 20% deviation from the ANPR travel time). The percentage of trajectory days that meet this criterion is the final score for the supplier.

Calculation:

$$\text{MAPE} = \frac{100}{k} \times \Sigma_{k=1}^{K} \frac{|TT_k^t - TT_k^w|}{TT_k^w}$$

Where:

k: Symbol for delivery period

K: Total number of delivery periods

$TT_k{}^t$: Delivered actual FVD travel time for period k

$TT_k{}^w$: Average realized travel time for time period k - measured with a high-accuracy measurement system

Score:

MAPE overall score on travel times.

Minimum requirements: For 80% of the trajectories, the MAPE score should be less than 20%.

Appendix D: ROMO1 data paper

Vehicle Fleet Data for Cost Efficient Real-Time Road Surface Assessment

Jochen Hipp[1], Kay Massow[2], Rafael Grote[2], Siegmar Züfle[1], ggf. Jochen Rauh[1], Timo Espenschied[1], 1x Mercedes-Benz[1], Patrick Blume[1], Ilja Radusch[3], Bram Vreugdenhil[4], and N.N.[1]

[1] Mercedes-Benz AG, Böblingen, Germany
jochen.hipp@mercedes-benz.com,
[2] Daimler Center for Automotive Information Technology Innovations, Technische Universität Berlin, Berlin, Germany
kay.massow@dcaiti.com
[3] Fraunhofer Institute for Open Communication Systems (FOKUS), Berlin, Germany
ilja.radusch@fokus.fraunhofer.de
[4] Rijkswaterstaat, Delft, Netherlands
mail@rijk.......com

Abstract. Aus der Intro hier hochnehmen The abstract should summarize the contents of the paper using at least 70 and at most 150 words. It will be set in 9-point font size and be inset 1.0 cm from the right and left margins. There will be two blank lines before and after the Abstract. ...The abstract should summarize the contents of the paper using at least 70 and at most 150 words. It will be set in 9-point font size and be inset 1.0 cm from the right and left margins. There will be two blank lines before and after the Abstract. ...The abstract should summarize the contents of the paper using at least 70 and at most 150 words. It will be set in 9-point font size and be inset 1.0 cm from the right and left margins. There will be two blank lines before and after the Abstract. ...

Keywords: Road Roughness, Road Waviness, ISO 8608, IRI, Field Data, Probe Data, Big Data

1 Introduction

Obviously, road waviness is crucial for vehicle safety, service life of vehicle components and last but not least for driving comfort. This is especially true for highways where vehicles drive at high speeds. In addition timely road maintenance helps to reduce the overall costs as surface damages progress is more than linear in time.

Consequently, road maintainer, be it private companies or government agencies, spend a remarkable amount of budget, effort and time to monitor the status of their roads. Well-known approaches are assessing the road surface via road roughness classes according to ISO 8608 [1] and calculating the so-called IRI, the international roughness index [2]. Both have been very well established for many

years now, have their pros and cons, and their relationship has been studied, see e.g. [3]. In either case, elaborate and costly high precision road profiles for thousands of miles of road surfaces are necessary. Moreover, those measurements need to be repeated on a regular basis and require organizing measurement campaigns with dedicated and costly special-purpose vehicles.

Simultaneously, today's passenger cars are increasingly equipped with a broad range of sensors, impressive network and computing power and connectivity abilities. This development is driven especially by the needs of driver assistance systems and the requirements of automated driving.

In this paper, we show how to employ the capabilities of today's passenger cars to assess road roughness bypassing the tedious and costly measurement of a fine granular longitudinal road height profile. By analysing the values from the vehicle level sensors we derive estimations of roughness classes according to ISO 8608 and IRI-values for street surfaces. These sensors continuously measure the vertical movement of all four wheels with respect to the chassis. Obviously, this data contains a mixture of both the actual road height profile itself but also the vehicle dependent chassis oscillation.

Our paper is organized as follows: In section 2 we give a short overview on related work and briefly discuss the general drawbacks of today's approaches to use mobile sensors to assess road roughness. Then in section 3 we suggest transferring the absolute values from the vehicle level sensors from the time domain into the location domain via the exact vehicle speed. We call this first canonical approximation the pseudo road height profile. Now for frequency analysis as suggested by ISO 8608 we show that despite the vehicle's natural oscillation we can always identify an area in the power spectrum where only the road roughness is visible and no chassis movements at all. We show the effectiveness of our approach for several test tracks on public roads and on our testing ground. Nevertheless, in section 4 we argue for pragmatic reasons to choose IRI for our very first field study. We show that similarly to our approach described in section 3, that follows the idea of ISO 8608, appropriate data preparation allows us to calculate IRI-values from the vehicle level sensors. After these theoretical investigations, in section 5 we show the applicability of our approach in practice in a broad field study. Our investigation covers the entire highway and highway-like network of the Netherlands and it is based on about eight month of customer data for this area of approximately $42,000\text{km}^2$. We carry out a thorough comparison of our vehicle data-based road roughness estimation against ground truth data from traditional measurement. Beside analytical issues like GPS-accuracy or sample frequencies, practical application also comes with computational challenges in face of the sheer amount of available data. For some road segments we aggregate several thousands of transits per day. In addition data protection and anonymization must be guaranteed. The encouraging result is that road roughness based on data from serious-production vehicles is surprisingly accurate and due to its low costs and continuous availability is a highly valuable supplement to conventional road roughness assessments. We close in section 6 with an outlook on what are the apparent next steps in order to further improve our analytic

approaches, our general analysis setup to unlock the potential of vehicle data based measurements for the process of road maintenance.

2 Related Work

Road profile estimation relies on different methodologies to derive the profile of the road pavement. These methods can be categorized into three distinct types: contact measurement, non-contact measurement, and system response-based estimation.[4] Based on the derived road profile, typically the roughness of the road is characterized according to ISO 8608, IRI or other approaches [1–4].

Contact measurements employ sensors which obtain data through direct contact to the pavement. These sensors include the profilograph, a static version of the straightedge mounted on a wheel with a sensor that can measure deviations in height. Non-contact measurements estimate the road elevation by scanning the road surface with non-tactile sensors, such as cameras, laser, or radar. The third category refers to response-based methods that measure displacements and accelerations while passing over the road surface. Traditional measuring setups typically combine multiple non-tactile and displacement measurement techniques, e.g. [5, 6]. They rely on an elaborate and costly sensor systems that demand appropriate maintenance. In addition, measurement campaigns must be thoroughly planned and executed to pass each road surface at least every couple of years.

At the same time, mobile devices have become ubiquitous in our daily lives. Today's smartphones are equipped with multiple high precision sensors for displacements and accelerations, cameras, GNSS, microphone, etc. These devices offer impressive computing power and can execute complex data analysis tasks. Accordingly so-called Probe Data Performance Management (PDPM) emerged as a new discipline [7, 8]. There are already established commercial companies, e.g. [21, 22] that beside other analytics tasks also offer road surface assessment based on mobile devices.

Whereas the results we see are very promising from a technical point of view, there are several fundamental drawbacks that make practical application of mobile customer devices somehow difficult: First, a dedicated app must be installed on the smartphone. Second, the user must explicitly run this app whenever a measurement campaign needs to be started. Third, the mobile phone must not stay in the owner's pocket but typically needs to be fixed for example on the center console of the car to measure accelerations and displacements [7] or even worse needs to be installed in a certain position behind the front window to make use of it's camera. From our point of view this makes approaches based on smartphones insufficient for everyday use by ordinary drivers. Accordingly, our impression is that commercial providers aim at dedicated vehicle fleets like garbage collection vehicles or public transport and rely on instructed drivers.

In order to get around such restrictions modern series vehicles themselves provide a promising alternative for data collection. Especially premium vehicles with advanced driver assistance systems (ADAS) or implementing the first steps

in direction of fully autonomous driving, e.g. [9, 10], are equipped with sensors like mono and stereo cameras, ultra sonic, radar, lidar and differential GNSS. Further, the vehicle ego-motion is essential for the implementation of such ADAS functions. This implies, high precision accelerations, yaw rates, etc. for all three dimensions are commonly available [11]. In addiation computing power and high bandwidth communication channels both inside the vehicle and via mobile network to an external central server are also mandatory [12]. In the context of road roughness assessment, especially the above mentioned vehicle level sensors as part of advanced damper systems are relevant.

advanced damper systems?

Finally, what makes employing vehicles as mobile sensors most compelling, is that they do not suffer from the drawbacks of mobile devices mentioned above: The software is installed, maintained and run by the OEM, the sensors are always in place and calibrated. At the same time data protection and data privacy must be ensured. Among other, the latter implies customer consent, full transparency and the customer's full control on what the data is used for and no persistence of raw data. At Mercedes-Benz this is always guaranteed [19].

Nur Mercedes-Benz schreiben, wenn direkter Verweis auf Privacy Center möglich

3 Road Roughness According to ISO 8608 based on Vehicle Data

The longitudinal unevenness of a road surface $h(l)$ can be understood as the vertical displacement h along the distance l of a track. For assessment of such a road profile the ISO 8608 [1] proposes employing a PSD of the vertical displacement. The dimensions are m^3 versus rad/m angular spatial frequency, respectively m for the wavelength, and both axes are logarithmically scaled, see examples in Fig. 1.

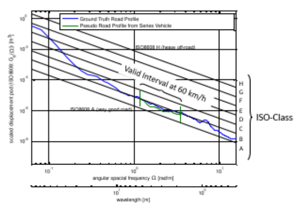

Fig. 1. Displacement Power Spectrum Densities for a Rural Road

In the next step ISO 8608 advises to smooth the PSD and to fit it by a straight line with the general form of

TNX RG!!!

$$G_d(\Omega) = G_d(\Omega_0) \cdot (\tfrac{\Omega}{\Omega_0})^{-w} \tag{1}$$

where $\Omega_0 = 1$ [rad/m] is the reference angular spatial frequency. Ω_0 serves as an anchor point and defines the height of the fitted PSD and w as exponent expresses the slope of the line. The height at Ω_0 can be interpreted as the intensity of the waviness. The value w expresses whether longer wavelengths (steeper line) or shorter wavelengths (flatter line) dominate the waviness of the road. As a further simplification ISO 8608 suggests assuming an exponent $w = 2$ and then assign ordinal roughness classes $A, B, C, \ldots, H$ based on $G_d(\Omega_0)$ to assess the road profile. Due to exponential limits of the roughness classes the distances in the sequence of the classes become linear. See ISO 8608 [1] for the actual calculation specification. The logarithm of $G_d(\Omega_0)$ corresponding to the limits of the roughness classes is also called AUN and is a linear continuous measure for road roughness. It supplements the roughness classes.

Rafael Grote: Wie berechnen wir den Wert? Das sollte hier hin

vermutlich Allgemeine Unebenheit, wo kommt das her?

Whereas this approach is commonly accepted and widely employed for road assessment it relies on an exact height profile of the road surface. At first sight the above-mentioned vehicle level sensors will hardly allow us to derive such a profile. Besides typical vehicle architecture related challenges like generally restricted measurement frequencies or insufficient resolution of GNSS positions, the vehicle itself is the reason: In the values from the vehicle level sensors, we obviously see a mixture of chassis excitation, wheel excitation, dampening of the tires together with the road waviness that we are actually interested in. Whereas very elaborate active vehicle suspension systems include acceleration sensors that would allow us to calculate out these effects [13], in practice this is not a feasible way to go. Vehicles with such a sensor setup would be comparably rare in customer fleets. In addition, appropriate and resource intensive analysis software in the chassis controller would be necessary.

The solution we propose is different: We do not try to separate vehicle excitation from road profile by back calculation. Instead, we make use of the known eigenfrequencies of the vehicle and its components, and the characteristics of the PSD fitted according to ISO 8608. Our idea is that there are frequency intervals where neither low frequent chassis movement nor high frequent wheel vibrations are visible. Then we consider the part of the spectrum where we can be sure that only the street waves are present. Presuming $w = 2$ from ISO 8608 allows us to derive $G_d(\Omega_0)$ from equation (1) with sufficient accuracy as we will show later. In detail our approach is as follows:

Ist w=2 nicht nur für die Einteilung in ISO-Klassen vorausgesetzt? Beim Fitten der Gerade kann sich doch auch eine andere Steigung (-w) ergeben. Wir ermitteln neben AUN auch w (Welligkeit)

1. Generate a so-called pseudo road profile from the vehicle level sensors. This is done by simply transferring the vehicle level signal as pseudo height from the time domain to the location domain via wheel RPM multiplied with the circumference of the wheel. Summing up the one-dimensional location in such a way presumes a very precise high-resolution wheel tick signal. GNSS as an alternative in our experience is much too inaccurate.

2. Thoroughly re-sample the pseudo profile to obtain equidistant values. This important step is further described in section 4.
3. Analyze the pseudo profile in the same way as a true road profile would be analyzed: We calculate a smoothed PSD on the vertical displacement as proposed in ISO 8608. Obviously, the resulting PSD-chart is not directly usable in the sense of the ISO because it contains both road waviness and excitation of the vehicle chassis and the wheels.
4. Calculate the limits of the interval in the wavelength domain that resembles pure road waviness in the spectrum. For that purpose we simply multiply the respective eigenfrequency with the current speed of the vehicle. 2Hz as lower bound from the chassis and 10Hz as upper bound for the wheels are gave us the best results. For a speed of 60km/h this approximately leads to the wavelength interval 8.3m - 1.6m, where only the road surface is present, see the limits drawn into Fig 1.

@Rafael Grote: Stimmt?

5. Now, as proposed in ISO 8608, we fit a straight line onto the smoothed PSD. As suggested in the ISO we assume w = 2 and only take the valid interval determined in the previous step into account for fitting the line.
6. From the fitted line we derive the value for $G_d(\Omega_0)$. Note that for that purpose Ω_0 needs not necessarily be part of the valid interval determined in step 3. Depending on the driven speed this won't be the case most of the time. Anyway, in that case the fitted straight line from step 4 is sufficient to determine $G_d(\Omega_0)$.
7. Finally, from $G_d(\Omega_0)$ we look up the roughness class according to the ISO specification and in addition we derive the supplementary AUN-value.

= Gd(?0) im linearisierten Spektrum

We implemented the approach from above in Matlab and validated it as follows: For about 15 road segments of a length of several hundred meters each we had high precision road profiles available as ground truth (stored as curved regular grid [20]. The tracks are public roads and roads on our testing ground. For those locations we collected wheel RPM and vehicle level sensor data in dedicated test drives. We did this by logging the vehicle bus data in a series vehicle at 50Hz during several passes per location. Then we derived the roughness class, respectively calculated the AUN-values for each track both from the ground truth and all vehicle drives separately. The result is shown in Fig. 2.

The outcomes based on the vehicle drives are plotted against the results based on ground truth data ("laser data"). The relevant ISO 8608 classes are shown as colored squares. We see a high correspondence between the results. There are only a view outliers where the vehicle data is still maximally one ISO class away from the ground truth. These results were also confirmed by simulations. By no means is this a comprehensive validation but from our point of view it shows the approach is worth testing in practical application.

RG: A few outliers? -¿ JH: Das S hat gefehlt?

4 Transfer to the Real World via IRI

For our practical investigation we have access to a large fleet of customer vehicles. All vehicles of this fleet are wirelessly connected to a backend server. From

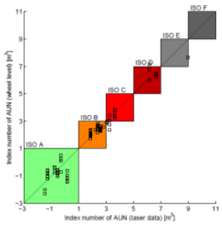

Fig. 2. Roughness classes from vehicle data versus ground truth ("laser data")

time to time the vehicles send short time series of selected sensor signals to the backend.[14] Due to the large number of vehicles the only serious limits for data availability are the costs for mobile data transmission, for computing power inside and outside the car and storage of intermediate results. The sheer number of vehicles allows us to generously reject every time series that does not fully match our expectations without losing overall coverage. Reasons for refusal are any signs of hard breaking, strong vehicle acceleration, strong lateral displacement, tire slip,, outliers, etc., or simply that the time series is rather short.

At the same time our goal is to push as much of the analysis task onto the vehicles themselves, like deriving the pseudo road surface profile. We do this to save bandwidth and to keep the data inside the car. For the vehicle level values we are currently restricted to 25Hz recording frequency for technical reasons. Mapping the vehicle level signal as pseudo height from the time domain to the location domain via interpolated GNSS positions turned out to be too inaccurate. Fortunately, inside the car the so-called wheel tics in conjunction with the wheel circumference allow a very precise estimation of the distance covered between the captured vehicle level values. Obviously, the distances between the height values then are no longer equidistant. Due to this and in order to obtain reproducibility over all pseudo profiles we re-sample the data with a high longitudinal resolution of 0.1m. For interpolation we employ the Akima spline algorithm [?], which combines the advantages of spline and linear interpolation (no overshooting). The resulting interpolated data is also augmented with GNSS coordinates. The latter are only be used for spatial aggregations in a later step.

Based on this real-world data we made a first try with the approach introduced above. In practice things turned out somehow more difficult than we expected. In brief we encountered the following obstacles:

Haben wir ein Bild "AUN aus Kundendaten"? GR: Ja, im Bereich Baden-Württemberg gibt es ein QGIS. Ist aber nur aus einem kurzen Zeitraum und sieht auf den ersten Blick nicht so toll aus. Ich schaue noch mal, ob wir noch was besseres haben.

1. For seriously calculating a PSD, road segments should be at least several hundred meters long. This was no issue for the given road segments during our tests where we manually chose beginning and end. Scaling to real world implies we need to automatize this step. For that purpose, we decided to rely on a map-based approach, i.e., we assign roughness values to edges from a common navigation map (SD-map), typically used by vehicle navigation systems and also implemented by OpenStreetMap [17]. Consequently, the segments are defined by the road network and for example in urban environments segments will often be much shorter than needed to calculate a meaningful PSD.
2. The valid interval taken to derive the smoothed PSD (step 3 above, previous section) is speed dependent. Whereas we do not see a severe impact of varying vehicle speed on the PSD itself, obviously the width of the valid interval shrinks if we take the minimal and maximal driven speed to constrain its borders.
3. The computational effort for the PSD calculation turns out to be tremendous due to the millions of time series we have available.

In a first attempt to come across those points we implemented an approach based on overlapping windows in order to obtain a fine granularity and at the same time run the PSD on sections of more than 100m length. We further filter our data to get only time series with rather constant speed. A way to cope with computational effort is to stop further analysis for a location as soon as a dozen overpasses are analyzed.

Whereas we see first promising results from above approach on customer data, and we are in general still convinced of the PSD-based approach, we learned that IRI for road waviness [2] is a serious alternative in practice. In brief Rijkswaterstaat and other relevant road infrastructure responsibles at least partly base their road maintenance work on regular IRI measurements. Moreover, at Rijkswaterstaat for several thousand miles of highway or highway like roads ("autosnelwegen") IRI values are available as ground truth for evaluation purposes. At the same time, we learned IRI calculation is not sensitive concerning above's points 1-3: time series sections can be rather short, speed differences are less impeding and for our scale up experiment most important, computational expense is only a fraction compared to our adapted version of waviness according to ISO 8608. The first steps from the above approach remain the same but instead of a frequency analysis we run the quarter car simulation from [2] on our transformed and re-sampled pseudo profile data. We tested different frequency filters to improve our results but did not yet get a clear picture on how this leads to general improvements. At the same time, we realized that already the canonical application shows acceptable results as we will show in the next section.

anders einleiten

5 Field Study

We conducted our field study in the Netherlands, focusing on over 5,000km of mainly highway or highway like roads, see Fig 3. Our study relies on vehicle data

Fig. 3. Mainly highway and highway-like roads considered for the files test. Background Map Data from OpenStreetMap

over for a period of eight months, from summer 2022 to spring 2023. Although we further restricted the data by collecting sensor values only for relevant roads in the area, we had to process a vast amount of data. The data stream can easily consist of about xy Gigabyte raw data per hour, depending on the time of the day. Data processing and aggregation of the results occurs on a daily basis on the time series data from the past hours. Given the sheer volume of data, Apache Spark is employed for the entire chain of processing as a highly scalable solution. By leveraging the efficiency-optimized functions within Spark, we accomplish the majority of the processing tasks, e.g. filtering, transformation, map-matching and aggregations. However, for more specific parts of the algorithms, e.g., resampling and interpolation as well as PSD- and IRI-calculation for the individual time series, we rely on vectorized data-processing techniques using Spark's user defined functions on Pandas data frames. This hybrid approach allows us to conduct effective and streamlined data analysis as follows: As a preparation step we split all edges of the underlying map into 10m segments. If an edge or its remaining part is too short we directly take it as a segment, no regard to how much shorter it is than 10m. We ended up with roughly the expected 500,000 map segments for the considered roads in the Netherlands.

Haben wir noch eine Zahl?

Stimmmt?

After data preprocessing, i.e. resampling and filtering out suspicious data, we also split all individual time series into 10m-sections. For each section we calculate the IRI-value according to [2]. Then we perform a spatial aggregation process. This involves determining the location of each time series section using

GNSS positions and matching it onto the corresponding edge segment on the map. Our approach is independent of any specific map as long it is modelled by edges and geolocated nodes, i.e. like a common SD-map is. By keeping the time series sections and the map segments with 10m rather short, we achieve more precise results during the aggregation of the IRI from the individual time series. Additionally, later this enables us to conduct a quite detailed comparison with the ground truth data. It's worth noting that due to our use of IRI as the roughness metric, we faced almost no limit on the minimum length of evaluated time series, unlike PSD-based techniques.

To validate our results we had access to a huge dataset of IRI-10 (10m road segment length) and IRI-100 (100m road segment length) values. This data, provided by Rijkswaterstaat, originates mainly from early 2022. For localization road geometries and geographic coordinates are included. This allows us to conduct a spatial matching process between the ground truth IRI and the IRI derived from the vehicle data.

For or comparison we decided to focus on IRI-10 instead of the more common IRI-100. This higher resolution implies less averaging when aggregating onto the map segments. Moreover, it enables us to be more fine granular when comparing the vehicle derived IRI-segments to the corresponding ground truth road segments. This is especially true when manually inspecting the results based on aerial images. Such visual inspection already revealed a strong coherence between the IRI-10 from our vehicles and the actual road roughness from the IRI-10 ground truth. For example, our approach is particularly effective in identifying railroad crossings, see Fig. 4, or the expansion joints at bridges. For visualization

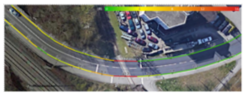

Fig. 4. Railroad crossing in Böblingen (Germany) with highly increased IRI-10 values measured by vehicle sensors (not part of field study area). Images ©2023 Google, GeoBasis-DE/BKG

we employed QGIS [16] with OSM [17] and aerial pictures as background.

The chosen validation area, consisting primarily of motorways known for their good condition in the Netherlands, exhibits the expected low IRI-10 ground truth values, typically clustered around 1m/km. Interestingly, we observed a very similar distribution for the IRI-10 values derived from the vehicle sensor data for the same area. Fig. 5 illustrates the IRI-10 distributions of both data sets and shows their congruence.

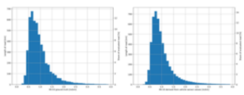

Fig. 5. IRI Distribution: Ground truth versus vehicle derived IRI

To take our evaluation further for each IRI-10 value from vehicle data we calculate its difference to the corresponding ground truth value. For that purpose we rely on matching the nearest geolocation. The distribution of these absolute differences highlights the level of agreement between the two data sets. At the same time we calculate the relative differences by dividing the absolute difference by the ground truth value. Both distributions are shown in Fig. 6.

Machen wir das so?

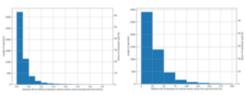

Fig. 6. Railroad crossing in Böblingen (Germany) with highly increased IRI-10 values measured by vehicle sensors (not part of field study area). Images ©2023 Google, GeoBasis-DE/BKG

The analysis confirms a strong coherence between the vehicle sensor-derived values and the ground truth data. The absolute difference between the two reveals a mean of 0.29m/km and a median of 0.18m/km. Similarly, the relative difference shows a mean of 31% deviation and a median of 21%.

Further examination of the data demonstrates the level of agreement between the vehicle sensor-derived IRI and the ground truth. A significant proportion of map segments, namely 63.0%, exhibit an absolute IRI deviation of 0.25m/km or less and 85.3% of map segments have an IRI deviation of 0.5m/km or less. Moreover, only a small percentage of 4.5% of map segments show an IRI deviation higher than 1.0m/km. We further examined these instances to determine whether they can be explained by specific road conditions, potential limitations in our methodology, or other factors contributing to variations in the roughness measurements.

It turned out the primary factor contributing to high IRI deviations was

ersten Satz rauslassen?

incorrect spatial alignment between the ground truth geometries and the mapped vehicle data. One key insight is that while the ground truth is available for individual lanes, our vehicle data was mapped to the single lane from the SD-map. Moreover, discrepancies in road geometries between the two data sets were observed. Additionally, weaknesses were identified in the spatial alignment of both maps, particularly in terms of lateral matching.

Although there is still a strong visual correlation in many cases (Fig. 12), improvements can be made to enhance the spatial matching between the maps. This will reduce the number of road segments to which a high deviation is assigned unjustified.

Another important reason for high deviations are actual changes in the road infrastructure itself. The time interval of several months between recording the ground truth and capturing the vehicle data seems relatively short to observe a general degradation of the roads. Nevertheless, significant variances show up in isolated cases. Many of those can be attributed to actual road damages, road works (both resulting in increased IRI), or completed maintenance activities (resulting in decreased IRI). In Fig.7 an example of a highway section that

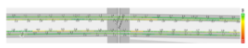

Fig. 7. Example for largely decreased IRI (upper line) after road maintenance (lower line). Images ©2023 Google, GeoBasis-DE/BKG

underwent maintenance, resulting in substantial IRI-10 improvements in that specific area is shown.

Dreased from x to y for the n m road

Furthermore, we investigated the influence of curves, deceleration, and acceleration as these directly have an effect on the vehicle level sensors. However, our analysis did not reveal any significant impact on the derived IRI-10 or its deviation from the ground truth.

Overall, the findings confirm the robustness of the vehicle sensor-derived IRI values and their close alignment with the ground truth data, indicating the reliability and effectiveness of the data collection and analysis methodology employed in the study.

Apart from the precision of the results themselves, we learned we are very fast, i.e., on frequented roads we know the IRI some hours after starting the data collection or even faster. Moreover, despite covering even large geographical areas (as mentioned above roughly 500,000 10m-edges), computational effort of IRI calculation turned out to be feasible. Aggregation of the daily data within this area takes about one hour on eight Spark nodes with 64GB RAM and 8 cores each. In addition, for performance evaluation we selected 50 million customer driven kilometers on a certain day in Europe and analysis together with spatial aggregation only took about 4 hours using the afore-mentioned Spark setup.

6 Conclusion and Future Work

In this paper we proposed a new approach to derive road roughness according to ISO 8608 and the IRI – international roughness index. Considering only series vehicles implies that no special data collection vehicles, no special hardware, system setup or maintenance of mobile devices are needed. Moreover, even dedicated measurement campaigns are not necessary and no instructed drivers are needed. Despite the obvious shortcomings of the sensor setup we are able to derive road roughness values that are promising to support the road maintenance process. In a broad evaluation on more than 5,000 road kilometers and Millions of analysed vehicle kilometers we showed the potential of our vehicle based IRI. Our IRI-10 values deviate only 0.29m/km in the average, respectively with a median of 0.18m/km, from ground truth data collected by dedicated measurement vehicle campaigns. Moreover, we are very fast. For main roads we obtain results in just some hours or even faster.

Nevertheless, we are aware that our work is only a very first step. We see several starting points to work on our IRI appoach, most important improving the map matching and distinguishing multiple lanes. Also we want to consider recent developments of improved IRI calculation [18]. Most important, for our evaluation we relied on IRI for pragmatic reasons only. We postponed to thoroughly evaluate our PSD-based approach following ISO 8608. Despite obvious challenges like length of segments and the imense computational effort we still believe the ISO roughness classes will be a valuable extension beyond the IRI.

Summing up, we are convinced that our approach is an extremely valuable supplement to make road assessment much cheaper, faster and more efficient. Despite all open technical questions, the key to raise this potential is a thorough integration into existing road maintenance processes.

References

1. ISO 8608: Mechanical Vibration – Road Surface Profiles – Reporting of Measured Data, International Standardization Organization, Geneva, Switzerland, 1995
2. Sayers, M. W.: On the Calculation of International Roughness Index from Longitudinal Road Profile In: TRANSPORTATION RESEARCH RECORD 1 50 (1995)
3. Kropáč, O., Múčka, P.: Be careful when using the International Roughness Index as an indicator of road unevenness, Journal of Sound and Vibration 287(4):989-1003 DOI:10.1016/j.jsv.2005.02.015
4. Nguyen, T., Lechner, B., Wong, Y. D.: Response-based methods to measure road surface irregularity: a state-of-the-art review, European Transport Research Review, 2019, https://doi.org/10.1186/s12544-019-0380-6
5. Fraunhofer MUM, https://www.ipm.fraunhofer.de/content/dam/ipm/en/PDFs/product-information/OF/MTS/mobile-urban-mapping-system-MUM.pdf RICHTIG ZITIEREN
6. Irgendwie die Autos vom Graefe zitieren? Gibt?s ein Paper? https://www.3d-mapping.de/
7. Sadjadi, S. Y.: Investigating for Road Roughness using Smartphone Sensors, International Journal Of Computers and Communications, Volume 11, 2017

8. Kumar, R., Mukherjee, A., Singh, V. P.: Community sensor network for monitoring road roughness using smartphones. Journal of Computing in Civil Engineering, Vol. 31, Issue 3, May 2017. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000624.
9. Ziegler, J. et. al.: Making bertha drive - an autonomous journey on a historic route, IEEE Intelligent Transportation Systems Magazine, 2014
10. FAP5-HAF
11. Marco, V. R., Kalkkuhl, J., Seel, T.: Nonlinear observer with observability-based parameter adaptation for vehicle motion estimation, IFAC Symposium on System Identification SYSID 2018 51.15, pages 60-65, 2018
12. Wirthmüller, F., Klimke, M., Schlechtriemen, J., Hipp, J., Reichert, M.: A fleet learning architecture for enhanced behavior predictions during challenging external conditions, IEEE Symposium Series on Computational Intelligence (SSCI), pages 2739-2745, 2020
13. Klinkner, W., Schmidt, F.; Apparatus for the road-surface-dependent control of shock absorbers in a vehicle suspension system, US Patent US4685698A, filed May 23, 1986
14. Wirthmüller, F., Klimke, M., Reichenbächer, C., Hipp, J., Reichert, M.: The Atlas of Lane Changes: Investigating Location-dependent Lane Change Behaviors Using Measurement Data from a Customer Fleet, IEEE International Intelligent Transportation Systems Conference (ITSC), pages 1225-1232, 2021
15. Massow, K., Kwella, B., Pfeifer, N., Häusler, F., Pontow, J., Radusch, I., Hipp, J., Dölitzscher, F., Haueis, M.: Deriving HD maps for highly automated driving from vehicular probe data, IEEE 19th International Conference on Intelligent Transportation Systems (ITSC), pages 1745-1752, 2016
16. QGIS: Welcome to the QGIS project!, https://www.qgis.org/en/site/, Date accessed June 15, 2023
17. OpenStreetMap, https://www.openstreetmap.org/, Date accessed June 15, 2023
18. Neue IRI-Berechnung
19.
20. ASAM: ASAM Open CRG, Version 1.2.0, 2020, https://www.asam.net/standards/detail/opencrg/
21. Argamon
22. Vialytics

Table 9 Summary of vehicle fleet data for cost efficient real-time road surface assessment

Overview	Details
Objective	Provide a cost-effective alternative for real-time road surface assessment using vehicle fleet data.
Methodology	Utilize vehicle sensor data to create a pseudo road profile.
	Apply frequency analysis to separate road surface characteristics from vehicle noise.
	Test with a fleet of vehicles across the Netherlands and compare with traditional methods.
Results	Vehicle-derived road roughness data closely matched traditional measurement methods.
	Demonstrated the ability to classify road surfaces accurately according to ISO 8608.
Conclusions	The approach offers a reliable, low-cost supplement to conventional road assessment methods.
	Future research will focus on enhancing analytical methods and exploring additional applications for road maintenance and safety.

Appendix E: Outcomes and lessons learned – GRM tables

To simplify navigation, we present the outcomes and lessons learned in GRM. The GRM outlines the consolidated key considerations of each phase, and the corresponding answers along with higher-level lessons learned. Additionally, the GRM indicates how to find the detailed answers and background information in this report.

The Outcomes and lessons learned column starts with the general lessons learned and followed with the outcomes gathered from the workshops and data stream exploration. The Map column shows the detailed locations of key insights and takeaways from the workshops in Appendix A1, A2 and B.

By thoroughly documenting the shared experiences of the selected NRAs, Appendix E realise that the lessons learned are not only preserved but are also mapped into actionable advice for NRAs across Europe. The GRM provides a referenceable overview of the advice corresponding to use cases for NRAs to consult when embarking on third-party data procurement initiatives. It is a crucial step in advancing the PRESORT project's goal of empowering NRAs to efficiently acquire and utilize third-party data.

Table 10 Pre-procurement phase - Guidance Reference Map

Consolidated key considerations	**Map**	**Outcomes and lessons learned**
Pre-study: research, trial, initiative, POCs, pilot	Appendix A1 Row 3, 5 Appendix A2 Row 1 Appendix B Row 3, 11	Lessons learned: Pre-studies such as research, trials, initiatives, proof-of-concepts, and pilots are essential in the pre-procurement phase, to find the criteria that define the goals of third-party data procurement. These pre-studies are typically smaller in scale, wireframe-based, with limited scope and (descriptive) functionality. Their purpose is to provide insights for potential procurements, which may occur years later. *Outcomes:* *Before FVD procurement, Zuid-Holland trial was performed successfully. Other pre-study (trials) was also performed to test non-real-time data with data quality methods. Research was done to develop algorithms for travel time calculation.* *Before ROMO1 procurement, NL NRA had experience on getting probe vehicle data for safety related traffic information with car manufactures free of charge (stated in ITS directive). Procuring the data for road maintenance to promote incentives for collaboration. 4 or 5 small-scale POCs were performed.*

		IE NRA is also performing third-party data trials to investigate the benefits of utilising and integrating to traditional NRA data. *Pre-studies on eCall data demonstrated possible time savings on equipped motorways and even greater savings on less-equipped rural roads, proving its value to be seen as a forgotten traffic management data.*
Market consultation (MC)	Appendix A1 Row 4 Appendix A2 Row 14	Lessons learned: MC plays a vital role in helping NRAs gain valuable insights into the capabilities and offerings of third-party data providers. By engaging in MC, NRAs are better equipped to understand the latest technologies, market trends, and available competencies, which enables them to make more informed and strategic decisions during the procurement phase. This process ensures that NRAs can tailor their data needs effectively, select the most suitable providers, and optimize their procurement strategies to improve long-term outcomes. *Outcomes:* *NDW recognizes the importance of MC in improving their understanding of third-party data, technology, and providers. MC helps NDW make informed decisions, such as updating maps with third-party data while developing an in-house system for flexibility in switching providers. It also enables NDW to better define data needs and identify potential partners.* *In preparation of the ROMO2, a market consultation was raised to gain insights. The knowledge built up during this market consultation such as 1. state-of-the-art innovative measures to achieve better quality, more efficiency and less costs; 2. identifying complex third-party data utilisation in different use cases[18], were used to optimally formulate the ROMO2 procurement.*
How to determine the procurement goals	Appendix A1 Row 3 Appendix A2 Row 2	Lessons learned: Gathered from the current practice of NL NRA third-party data procurement during the workshops, the criteria to determine third-party data procurement goals are based on validating the effectiveness of data for these specific goals, considering both technological innovation and budget feasibility. For innovative solutions, these criteria shall embrace continuous improvement to enhance traffic and

[18] ROMO2 procurement was published online as of October 9th, 2024. The previous ROMO1 procurement and the market consultation collected knowledge to build the 12 use cases: *Disintegration of an asphalt road surface, Longitudinal unevenness (IRI100), Height difference longitudinal direction II, Road markings for asset management, Fuel consumption/energy consumption related to the road surface/conditions, Road friction information, Road temperatures, Air temperatures and other weather conditions, Air temperatures and other weather conditions, Decline in visibility of existing road markings, Safety hotspots, Road markings for safety purposes, Accidents and stationary vehicles.*

		asset management. *Outcomes:* *For FVD, pre-studies such as research and trials, developed the criteria to achieve travel time data calculation using non-real-time traffic data.* *For ROMO1, the procurement goal was to assess the validity of using probe vehicle data for asset management. Only non-quantifiable criteria exist to validate whether the procured data met the goal.*
Stakeholder engagement	Appendix A1 Row 16 Appendix A2 Row 6	Lessons learned: Stakeholder engagement is critical for effective traffic data management, ensuring collaboration and alignment between various parties involved in the pre-procurement, procurement, and implementation of data. Strong communication and cooperation between stakeholders help NRAs make better decisions and deliver high-quality services utilising third-party data. *Outcomes:* *NDW exemplifies stakeholder engagement through its collaboration with governmental partners, including IenW, RWS, provinces, municipalities, and city regions. This partnership allows NDW to gather traffic data from traditional loop detectors cameras, to procure third-party data from multiple sources (FVD, ROMO and many others), and to integrate and aggregate with a self-build in-house system. NDW then distributes some processed data to the public for free of use. It also provides its products to its partnership governmental organizations for free and to companies with licence fee. This can improve traffic management and policymaking. By working closely with these stakeholders, NDW ensures a coordinated approach to traffic data collection and distribution across a broad network.* *In ROMO1, IenW communicates regularly with car manufacturers (third-party data provider) to ensure ROMO1 project implementation and address any issues. This collaboration helps both parties stay aligned on technological updates and maintain on track to achieve ROMO1 goals.*

Table 11 Procurement phase - Guidance Reference Map

Consolidated key consideration	Map	Outcomes and lessons learned details

Rationale of procurement	Appendix A1 Row 1 Appendix A2 Row 1	Lessons learned: Taken from the workshops, third-party data procurement offers a strategic approach to overcoming limitations in traditional traffic monitoring systems. By leveraging external data sources, NRAs can enhance data coverage, improve data reliability, and reduce data collection costs. This approach supports more effective traffic management, road safety, and infrastructure planning, all while addressing budget constraints and the need for scalable, innovative solutions. *Outcomes:* *FVD procurement allows for broader travel time coverage at a lower cost compared to loop detection and camera systems, which are expensive and provide unstable data. FVD also offers greater flexibility and scalability for comprehensive traffic monitoring.* *ROMO1 procurement leverages probe vehicle in-vehicle data for road management, safety, and asset management. With budget constraints limiting physical infrastructure maintenance, ROMO1 explore third-party data to offer an innovative solution to asset management.*
Data requirements and criteria	Appendix A1 Row 7	Lessons learned: Setting **minimum requirements** for third-party data during the procurement phase is crucial to ensuring data quality and integrity. These requirements help guarantee that the data aligns with the organization's standards and fits seamlessly into existing systems. Moreover, establishing these minimum thresholds prevents low-quality datasets from being accepted, even if they come at a low price. This prevents cost-driven procurement from undermining data reliability and ensures that quality takes priority over cost. *Outcomes:* *In the case of FVD, NDW requires third-party data providers to meet certain minimum requirements, such as providing a real-time data feed for two weeks. The data is then compared against NDW's own ground truth using tools like speed heatmaps and methods like RMSE and MAPE to assess data accuracy. NDW ensures that data with lower quality can't be compensated by a lower price, thus maintaining high data standards throughout the procurement process.*
	Appendix A1 Row 8 Appendix A2 Row 4	Lessons learned: Establishing **award criteria** during the procurement phase is crucial for ensuring both high data quality and cost-effectiveness. These criteria provide a structured approach to evaluate bids, where the highest score wins, ensuring that bids are assessed based on a combination of quality and cost

		factors. This method ensures a balanced consideration of both data accuracy and the financial aspects of the bids. *Outcomes:* *In the FVD procurement by NDW, bids are evaluated based on a set of award criteria. These include factors like bid amount, data quality, and timeliness, using methods such as RMSE and MAPE for accuracy. Each criterion is scored, with the highest overall score winning the tender. This ensures that both quality and cost are fairly considered in the selection process, and that the bid with the best balance of these factors is awarded the contract.*
Procurement award models	Appendix A1 Row 8 Appendix A2 Row 3, 19	Lessons learned: Procurement award models offer different methods for balancing price and quality when selecting third-party data. The current common model is BPQR, which considers both cost and quality, ensuring a fair evaluation process. It is more comprehensive compared to models like Lowest Price, which focus solely on cost and risk sacrificing quality. BPQR is widely regarded as a more balanced approach, ensuring high-quality and maintaining competitive pricing. *Outcomes:* *For example, in the FVD procurement by NDW, BPQR and scoring system rewards bidders that offer the best overall value, rather than simply choosing the lowest price. This ensures high-quality data.* *ROMO1 also employs BPQR procurement model. Positioning its active involvement in the field, Mercedes-Benz secured all three lots in the tender by offering a strategic series of bids in which the low price and high quality of their bids really stood out. Unlike many car manufacturers that rely on other companies for data aggregation, Mercedes-Benz manages its vehicle data in-house. Through ROMO1, they repurpose previously underutilized data, such as missing lane information detected by Lane Keeping Systems, adding value and enhancing their competitive edge in the procurement process.* *In the ongoing ROMO2 procurement, an innovative procurement framework is being conceptualized, which incorporated variable renumeration as a key element. This approach allows for flexibility using volume quality matrix[19].*
Licensing and	Appendix A1	Lessons learned:

19 https://www.tenderned.nl/aankondigingen/overzicht/352683

usage	Row 10, 14	Data licensing in public procurement is vital for NRAs, as it determines who can access, share, and monetize the data. NRAs often seek agreements that prevent third-party data providers from commercially exploiting the data, ensuring it is primarily used for public interest. However, negotiations vary, and some NRAs may impose strict limits, such as prohibiting redistribution. At a minimum, NRAs require control over data usage and aim to balance public benefit with data security. *Outcomes:* *NDW offers a flexible licensing model. While government agencies access NDW data for free, non-governmental entities can access NDW products under a paid license. Furthermore, NDW makes certain aggregated data available to the public through open data platforms, encouraging innovation and transparency. This approach balances the open data principle with the need to monetize certain data for operational sustainability. The NDW's licensing scheme is planned during procurement, third-party data is procured under a license (meaning that FVD can only be made available to road authorities and for tasks assigned by road authorities). Moreover, procuring off-the-shelf data from third-party providers is also strategic to encourage broader access while still able to manage data ownership and usage rights.*

Table 12 Implementation phase - Guidance Reference Map

Consolidated Key considerations	**Map**	**Outcomes and lessons learned details**
Data attributes specifications	Appendix A1 Row 9 Appendix A2 Row 10	Lessons learned: Data attributes align with the requirements and criteria during the procurement phase and are critical in the implementation phase. Key data attributes like accuracy, coverage, and update frequency ensure seamless integration into existing systems and fulfilment of operational needs. Defining these attributes early guarantees reliable data delivery that supports real-time traffic management and decision-making. *Outcomes:* *For example, in NDW's procurement of FVD, key attributes such as segment size, coverage, and accuracy were essential. During implementation, NDW integrates a monthly basemap provided by the supplier and performs map-matching to ensure the data fits NDW's in-house systems, like the "Viewer" and "Trajectenmaker." This ensures NDW can maintain high accuracy and timely traffic*

		data for public and governmental use. *The implementation phase of ROMO1 required more technical data than initially anticipated, such as road segment information for slippery conditions and temperature readings. Originally, the focus was on descriptive and functionality of delivering a dashboard rather than technical data specs. The later needs underscored the importance of specifying these detailed attributes during procurement to support a smoother implementation.*
Data integration	Appendix A1 Row 6, 9, 11 Appendix A2 Row 9 Appendix B Row 8	Lessons learned: Effective data integration is essential for combining diverse data sources to meet complex project objectives. It ensures accuracy, consistency, and alignment with system requirements, enabling seamless data integration with existing data. Plans for how to integrate third-party data with existing ones (e.g., systems, APIs, databases) must be considered in previous phases before implementation phase to have strong project outcomes. *Outcomes:* *By acquiring off-the-shelf third-party data, NDW reduces costs and ensures transparency in how the data are generated, making it easier to integrate and aggregate with in-house systems and products. NDW's in-house knowledge and systems provide strong base for off-the-shelf data integration, allowing NDW to apply its own FVD algorithms for integration, such as aggregating and generating travel time over a trajectory. A dedicated team ensures that new maps are integrated monthly, updating and expanding the data's reach and accuracy as NDW plans to do over years.* *In ROMO1, data integration begins with a dashboard displaying visualized third-party data for events like potholes, de-icing etc., rather than directly integrating it into the NDW system. The dashboard acts as a second screen to show the events. Once the information is verified, the data is integrated into NDW's in-house systems and products, such as the "Viewer". Additionally, Noord Holland province has its own API connection with the dashboard, facilitating more localized data integration within its infrastructure.* *In the early trial phase of data integration for the Ireland NRA with Valerann, the process remains offline and separate from in-house databases. Consultants in the control centre monitors real-time events and incident alerts via camera feeds. While not yet implemented, the trials underscore the importance of developing filters for inaccurate data, highlighting the need for robust integration to ensure reliable road management.*
System	Appendix A1	Lessons learned:

operations and data extraction	Row 12 Appendix A2 Row 5	System operation and data extraction of procured data require ensuring compatibility with existing platforms and proper adaptation for smooth integration. Aligning with internal systems and needs, which involves significant adjustments, procured data is pulled or pushed. The data extraction is usually tailored for maintaining performance, accuracy, and compliance with data protection regulations. Early planning during procurement reduces the effort needed for adaptation and ensures data utility during implementation. *Among traffic safety and traffic management products, NDW created two in-house real-time systems to handle the data: the "Viewer" for real-time traffic visualization and "Trajectenmaker" for hosting specific trajectories. Data is pulled from the procured FVD, ensuring that targeted information is retrieved and integrates seamlessly into NDW's systems, enabling effective real-time monitoring.* *ROMO1 operation is complicated by data privacy regulations. While safety-related data is required to be shared, additional data requires drivers to opt-in, i.e. drivers opt in for sharing information with car manufacturers. Intended for asset and road management. Car manufacturers (third-party data providers) are tasked with anonymizing this information to ensure the strict protection of personal data under GDPR.*
Data access control	Appendix A1 Row 12 Appendix A2 Row 7,8,9	Lessons learned: Data access control for procured data and its derivatives must balance open access for public use and controlled distribution for private entities. Procured by governments, governmental organizations should be given free access to improve traffic management and public safety, while private companies may access it via paid licenses. This model promotes data usage of NRAs while ensuring that the data's value is preserved and incentivizes innovation in traffic solutions. *Outcomes:* *The FVD data was used to create various TM products, such as real-time traffic monitoring and congestion analysis tools. These products are freely available to governmental organizations through platforms like NDW's "Viewer," allowing them to monitor traffic conditions in real-time. Private companies, however, can access these products via a licensing model with extra conditions. This flexibility enables broader data usage while maintaining NDW's control over how the data is distributed, also discussed in Table 11.* *In ROMO1, a dashboard was initially used to visualize data and provide traffic insights before fully integrating it into NDW's "Viewer". Government organizations could freely access ROMO's derived road safety and asset management data, supporting more localized TM efforts. The dashboard facilitated the road monitoring of using large datasets, allowing operators with varying levels of data*

		expertise to identify safety hotspots and critical events.

Table 13 Maintenance and monitoring phase - Guidance Reference Map

Consolidated key considerations	**Map**	**Outcomes and lessons learned details**
Data quality maintenance monitoring	Appendix A1 Row 13	Lessons learned: Maintenance and monitoring must focus on maintaining high data quality throughout the operation. This includes regularly ensuring that metrics such as accuracy, coverage, and timeliness stay aligned with the KPIs outlined in the award criteria during the procurement phase. Regular monitoring ensures that the data continues to meet operational standards, enabling effective real-time traffic management and long-term system effectiveness. *Outcomes:* *In FVD, NDW handles maintenance in-house to ensure best practice of in-house systems. Every minute real-time data pulling (especially during peak traffic hours) was addressed such as slow data retrieval, a solution was implemented where data was pulled every 10 seconds and processed within 50 seconds, maintaining data integrity and timeliness as required during the procurement phase.*
NRA and third-party data provider communication	Appendix A2 Row 6,7,10	Lessons learned: Effective communication and regular collaboration ensure that stakeholders' evolving needs and issues are addressed and that the system remains responsive during operation, maintenance and monitoring. Establishing clear communication protocols facilitates smoother navigation of complex data pools and ensures that the system is adaptable to changing requirements. *Outcomes:* *In ROMO1, the procurement initially focused on descriptive outcomes and lacked sufficient technical details, resulting in missing data attributes such as road segment temperature and slipperiness. This gap, identified in this phase, highlighted the importance of continuous dialogue to refine the data attributes for better system performance.*
Monitoring	Appendix A1	Lessons learned:

protocols	Row 21	Monitoring protocols ensure that problems observed such as system disruptions or data inaccuracies, are identified and addressed promptly, maintaining overall system performance. *Outcomes:* *In FVD, quarterly reporting on maintenance and monitoring is conducted to track any issues related to uptime and downtime, ensuring they meet the required thresholds. Issues such as downtime exceeding the maximum allowance of 8 hours per year are escalated for quick resolution, maintaining system performance and service reliability.*

Table 14 SLA, scalability, challenges and risk - Guidance Reference Map

Consolidated key considerations	**Map**	**Outcomes and lessons learned details**
SLA monitoring	Appendix A1 Row 16,20,21 Appendix A2 Row 12	Lessons learned: SLA monitoring focuses on ensuring that key service metrics—such as data quality, uptime or downtime, latency, and response times—are continuously evaluated and meet the standards. Regular testing of the procured third-party data and reporting up-to date allow for early detection of issues, maintaining service reliability and operational efficiency. *Outcomes:* *In FVD, NDW conducts monthly data quality tests of procured data on road trajectories using the same criteria as defined in the tender. These tests, combined with quarterly SLA reporting, help monitor data delivery latency and uptime, ensuring that downtime stays within the required threshold to avoid disruptions.* *In ROMO1, SLA metrics also track the uptime and downtime of the dashboard and API response times, ensuring minimal data delivery latency and erroneous event detection.*
Scalability	Appendix A1 Row 13,15 Appendix A2 Row 1,12	Lessons learned: Scalability is one of the key considerations when expanding infrastructure coverage for data-driven road management. Traditional detection methods can be prohibitively expensive or unfeasible, especially for large-scale or real-time monitoring of road management. Scalable in-house data

	Appendix B Row 13	systems need to process large volumes of data efficiently and extend coverage beyond main roads to secondary or rural areas. *Outcomes:* *In FVD, FVD covered at first only regionally. With eventual procurements of FVD, more road network is covered with FVD with a cost-efficient price, since the main system set-up itself can be used for a bigger network. The systems can handle scale-up data volumes, like real-time updates, even as the geographical and/or road types coverage expand.* *For ROMO1, scalable use of probe vehicle data enhances asset and road management, enabling coverage expansion without costly physical infrastructure. With several small-scale pilots showing the benefits, the procured dashboard built upon probe-vehicle in-vehicle data helps improve road safety and TM.* *In eCall study, scalability plays a crucial role in improving emergency response systems. The next generation of eCall can potential be included to public safety answering points (PSAPS), to save time and improve emergency response efficiency on a broader scale.*
Barriers/ Challenges	Appendix A1 Row 15 Appendix A2 Row 13 Appendix B Row 14	Lessons learned: Challenges in third-party data procurement vary significantly, ranging from licensing restrictions to organizational complexities. Procurement challenges often stem from the upscaling of data management, internal resistance, high cost on in-house systems development, data collection/integration difficulty across systems and systems update due to change of third-party data vendors. *Outcomes:* *In FVD, challenges included a limited number of bids for off-the-shelf data, as companies preferred to develop their own products for future licensing profits. Additionally, segment data lacked critical Origin-Destination (OD) details, needing a separate procurement.* *In ROMO1, collaboration hurdles between car manufacturers and road operators highlighted the complexity of cross-organizational efforts. Road operators, not on the ROMO payroll but often volunteers, required a simple action list to address dashboard. This also coincides with* strict budget challenge on road infrastructure maintenance, which cause internal resistance to perform innovative project like ROMO. *In third-party data trials IE NRA, filtering erroneous vehicle data was a key challenge. Developing mechanisms to eliminate inaccuracies, such as filtering out stopped vehicles to improve third-party*

		data reliability.
Risk	Appendix A1 Row 14,15,17,18 Appendix A2 Row 17,18,19,20,21 Appendix B Row 16,17	Lessons learned: In third-party data procurement, both anticipated and unforeseen risks arise, such as cybersecurity, regulatory concerns, extra cost for system update etc. These risks need mitigation strategies to ensure the continued integrity and usefulness of the data while avoiding system disruptions. *Outcomes:* *FVD risks:* *- Massive data volume: Managing and processing vast datasets, where even a day with minute of data can exceed 10 GB, presents risks of delayed or impractical processing times, hindering real-time data integration.* *Mitigation: Scalable systems are developed in-house to handle such immense volumes, ensuring real-time functionality.* *- Data size management risk: Some providers reduce data size by removing headers, which can lead to problems during data integration and interpretation.* *Mitigation: Enforce strict data format requirements during procurement, ensuring necessary metadata remains intact for seamless integration.* *- Fleet stability risk: Dependence on a stable fleet to provide FVD data is crucial.* *Mitigation: Contracts and agreements with stable providers and fixed fleet requirement ensure continuous data streams, while exigent situations like provider bankruptcy are prepared for by including contingency clauses.* *- Policy and regulation instability: Sudden changes in policy or regulations could disrupt operations.* *Mitigation: Alignment with stable regulatory policies to avoid disruptions and include regular compliance checks in procurement criteria.* *- Change of provider risk: Switching data providers requires a significant system update due to change of algorithms and integration (i.e. map-matching) used by new providers.* *Mitigation: Develop adaptable in-house systems that minimize downtime during provider transitions, blocking the NRA development team's schedule to ensure system update, data compatibility and continuous operation.*

		ROMO1 risks: *- Vendor locking risk: A risk arose when one supplier won all three procurement lots at the same time, creating dependency on a single provider.* *Mitigation: Prevent alliances by encouraging multiple data sources and issuing separate tenders for 12 use case in ROMO2 to diversify providers.* *- Lack of standardisation: Road management is still low on standardisation, especially in innovative data exploitation.* *Mitigation: Collaborate with research institutions like TNO NL to research and promote standardisation.* *-Traditional procurement model risk: The traditional procurement seeking one supplier could lead to suboptimal outcomes.* *- Mitigation: Engage procurement lawyers to devise frameworks and explore variable renumeration based on data quality and volume.* *- Cybersecurity risk and GDPR: Ensuring data security and compliance with GDPR.* *Mitigation: Strict monitoring by car manufacturers and explicit inclusion of GDPR and cybersecurity clauses in procurement documentation.* *eCall risks:* *- Data anonymization: eCall data contains personal information, raising GDPR concerns.* *Mitigation: Implement GDPR-compliant data anonymization, leveraging the real-time nature of eCall data, which falls under more flexible regulations for emergency data.* *- Technology update risk: The transition from 2G and 3G networks, affecting the functionality of eCall systems.* *Mitigation: Future-proof by ensuring 4G and 5G compatibility in new vehicles while maintaining 2G where necessary for legacy systems (not just eCall system).*