Paper ID EU-TP1506

Towards a National Floating Car Data Platform for Austria

Karl Rehrl[1*], Stefan Henneberger[1], Sven Leitinger[1], Andreas Wagner[1], Michael Wimmer[1]

1. Salzburg Research Forschungsgesellschaft mbH, Austria, karl.rehrl@salzburgresearch.at

Abstract

Floating Car Data (FCD) are an essential data source for real-time traffic information services. While FCD processing is typically the expertise of commercial service providers, road authorities are supposed to buy processed traffic information for operating their own traffic information services. However, beside considerable costs, contracting vehicle-originated traffic information often limits road authorities' data usage rights. In Austria, the federal road authorities opted for an alternative way. As part of the national EVIS.AT platform for real-time road traffic information, the road authorities decided to establish a national FCD platform. The operator of the FCD platform is responsible for organizational as well as operational aspects concerning FCD collection, processing and sharing. This work describes the technical and data privacy concepts of the platform, the introduced quality measures and experiences from daily operation.

Keywords:

Floating Car Data, National Platform, Public Road Authorities

Introduction

Floating Car Data (FCD) are an essential data source for real-time traffic information services. A broad variety of nowadays services uses FCD for estimating real-time or historic travel times. Typically, commercial traffic information service providers (TISPs) such as TomTom, Here or INRIX operate FCD-enabled traffic information services and handle the complexity of FCD processing. For using FCD as source for their own traffic information or traffic management services, road authorities are supposed to contract commercial FCD-based traffic information services. Beside considerable costs, due to other commercial contracts, TISPs often limit usage rights especially concerning data storage and re-use as well as publication of derived traffic information products. Moreover, road authorities have to integrate the acquired data with their legacy systems, which often requires on-the-fly mapping to local road referencing systems as well as continuous quality assurance. An example for a national data platform supporting this role model is the Netherlands' National Data Warehouse (Viti, Hoogendoorn, Immers, Tampère, & Lanser, 2008). A different role model pursues the collection and processing of raw FCD (time-referenced anonymous vehicle trajectories) directly by the road authorities. This role model faces the drawback that road authorities have to operate a FCD system on their own fostering the need for considerable knowledge on FCD processing. Beside the technical aspects, road authorities are supposed to deal with a variety of organizational issues such as fleet contracting or data privacy. However, the benefit of such a role model is to gain full control over data processing from raw FCD to derived traffic information products including full data usage rights. Especially in case of using the data for traffic analytics and/or traffic management, this can be a significant advantage.

In Austria, the Federal Road Authorities (led by the national motorway operator ASFINAG and supported by the Federal Ministry of Transport, Innovation and Technology) have decided to implement the second role model within their national initiative to establish a platform for real-time road traffic information called EVIS.AT[1]. Goal of EVIS.AT is to generate and share harmonized road traffic information (planned and unplanned incidents, current and forecasted travel times) in near real-time between federal road authorities laying the foundation for public traffic information services. A sub-objective of the EVIS.AT project is to establish and operate a national FCD platform for collecting and processing FCD as well as sharing FCD-based, road-referenced travel times between road authorities. The intention of this national platform is to centralize the complex processing of FCD while serving the federal road authorities with quality-assured real-time travel time data for further processing. The overall concept of the national FCD platform has been designed, implemented and tested since 2012 within the nationally funded Austrian flagship project FCD Testbed Salzburg (Brunauer & Rehrl, 2014; Rehrl, Brunauer, & Gröchenig, 2016), which in the context of EVIS.AT has been extended to nation-wide coverage. By the end of 2017, the EVIS.AT FCD platform has been operational as pilot system processing nearly 30 million vehicle locations on weekdays from more than 5.500 connected vehicles during peak hours. The work describes technical concepts including the data processing pipeline, implemented data privacy concepts as well as implemented data quality monitoring concepts. Moreover, the work gives first insights on daily operational aspects, especially concerning FCD quality measures.

Technical concepts

Overall, the objective of the Austrian national FCD platform is to handle FCD data sources at one central point, either technically as well as organizationally. Organizationally means, that one organization within the EVIS.AT platform is responsible for FCD acquisition, negotiating and managing contracts with fleets, ensuring data privacy and realizing interfaces to telematics service providers. The organization operating the FCD platform is responsible for collecting the data from contracted fleets via different telematics service providers, taking care of data cleaning and anonymization as well as processing the FCD raw data to link-based travel times being referenced to the Austrian National Transport Reference System GIP.at[2]. Fully anonymous travel times attributed with metadata about fleet characteristics and quality parameters are shared via the EVIS.AT platform, which operates as central data hub for the federal road authorities.

[1] http://www.evis.gv.at

[2] http://gip.gv.at/home-en.html

System Architecture

The platform for collecting, analysing and sharing FCD provides four main processing steps:

1. **Collect**: The first processing step collects time-referenced and pre-anonymized GNSS measurements (e.g. coordinates from global navigation satellite systems - GNSS) from different data sources (e.g. fleet telematics systems, smartphones). This sub-system applies first data pre-processing (e.g. data validation and cleaning) and data transformation steps (e.g. transforming the data from proprietary data formats into a common format)
2. **Assemble**: The assembly step assembles the time-referenced GNSS measurements to time-ordered anonymous vehicle trajectories. These trajectories are attributed with metadata about fleet characteristics and in case of successful quality checks, are being transferred to the following processing pipeline. Anonymous vehicle trajectories are considered the basic data format for all further processing steps.
3. **Analyze & Match**: This step analyzes vehicle trajectories, identifies and cleans outliers, identifies typical motion patterns like single stops or stop and go traffic patterns and matches trajectories to the directed road links of the national transport graph. This processing step results in link-based travel times of single vehicles called 'travels' as well as additional traffic-relevant information patterns like vehicle stops or stop-and-go traffic being called 'motion patterns'. These motion patterns are referenced to the road links with relative offsets, meaning that they, for example, start at 25% of one link running until 57% of another link.
4. **Share**: The last step shares 'travels' and 'motion patterns' with the federal road authorities via the EVIS.AT real-time traffic information platform via a standardized data interface.

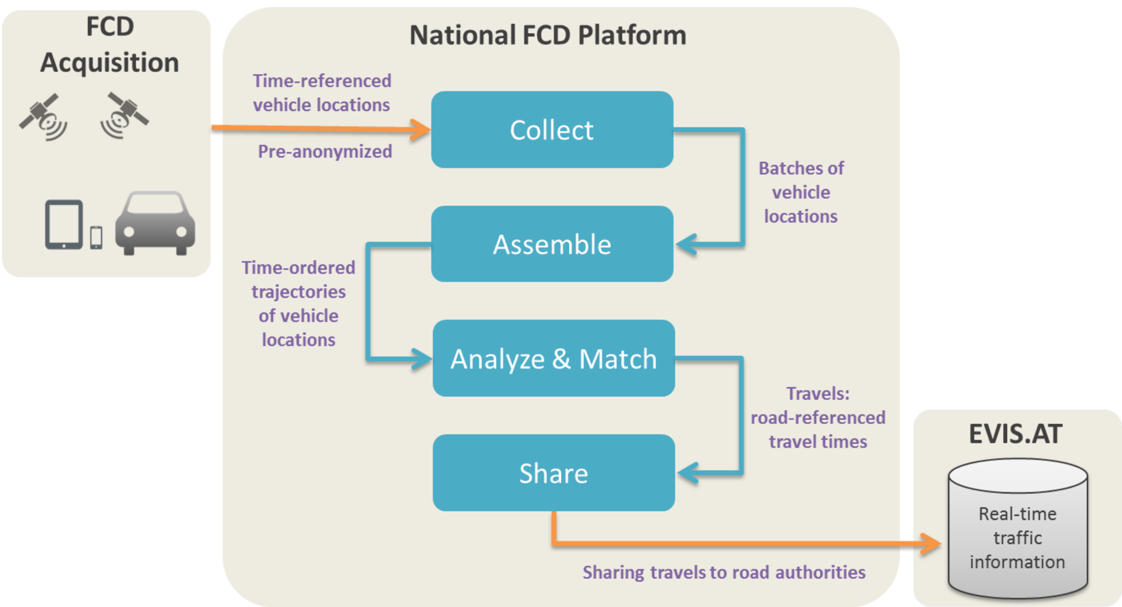

Figure 1 – System Architecture of the Austrian National FCD Platform

Collect

During the first processing step, time-referenced GNSS measurements from different data sources, mainly from telematics service providers, are collected via data interfaces (upon contracts with fleets). While most of the data is fetched from proprietary interfaces, the FCD platform also offers a standard interface for submitting time-referenced GNSS measurements. GNSS measurements are typically collected in batches, where each batch contains a randomly assigned trip ID being changed for every new trip. This means that the collected GNSS measurements are already pre-anonymized, ensuring that the data does not contain any attributes linked to persons or vehicle identifications. During the collection process, the data is harmonized to a standardized data format, which also includes first filtering and cleaning mechanisms (e.g. detecting and cleaning timestamp or coordinate duplicates or implausible coordinates). For further data processing, the data collection process forwards the cleaned raw GNSS measurements to the next sub-system.

Assemble

During the second step, the assembling process takes the pre-processed raw data and concatenates all GNSS measurements containing the same trip ID to time-ordered vehicle trajectories. If the trip ID has not previously been identified, a new (sub-) trajectory is started. After 10 minutes of not receiving new GNSS measurements, a trajectory is finished. This process also takes care of time-based ordering of GNSS measurements, ensuring that lately transmitted batches are correctly entered into the trajectory. (Sub-) trajectories of GNSS measurement are the basis for all further processing steps. Before the (sub-) trajectory is forwarded to the next sub-system for further processing, the assembler checks whether a trajectory already contains enough data (at least three GNSS measurements). Furthermore, continuous quality checks ensure that a certain level of quality is guaranteed at this stage. Based on metadata for data sources, fleet characteristics like heavy freight vehicles, public transport lines with bus lane permissions or ambulance vehicles and access rights are added as flags to the trajectory.

Analyze and Match

Each single trajectory is analyzed and matched separately and iteratively. Firstly, a pre-processing step filters and cleans the data (e.g. removing GNSS outliers) and applies validation procedures (e.g. checking the data against speed thresholds). The filtered, cleaned and validated (sub-) trajectory is submitted to the map matching process, which determines the most likely vehicle path with respect to the underlying road network. Map matching is accomplished with our own highly scalable map matching implementation being capable of matching only the new parts of a growing trajectory by extending the matched path of the last map matching iteration (Rehrl, Gröchenig, & Wimmer, 2018). Recently, the map matching implementation has been published as open source implementation as part of the Graphium project[3]. Next, the map matching result (the most likely vehicle path) is used to calculate link-based travel times of single vehicles for each directed link along the path using again the

---

[3] https://github.com/graphium-project/graphium-neo4j

GNSS measurements from the trajectory. For this calculation, entry and exit times of vehicles on road links are determined using geometrical linear interpolation (Thill, 2000) between GNSS measurements (we call the resulting link and direction-based travel times simply 'travels'). Using the road geometry also leads to average vehicle speeds. Travels together with FCD quality parameters (e.g. average GNSS sampling intervals, average map matching error, path validity) are shared with the federal road authorities in the next processing step.

It has to be mentioned that in addition to the basic parameters, the FCD platform also calculates enhanced traffic flow information called 'motion patterns', predominately by means of time-series analysis of single vehicle trajectories. The idea of motion patterns is to attribute travels with additional flow information, which may not be reconstructed from entry-to-exit travel times. Examples for motion patterns are single stops or stop and go traffic patterns. Motion patterns are linearly referenced in travel direction on road links and are shared with the federal road authorities as well.

Share

The main objective of the FCD platform is to share vehicle-originated travel time data with the federal road authorities being connected to the EVIS.AT platform (by transmitting new travels every 60 seconds). In order to serve the needs of the road authorities, the sharing service supports filtering by road network categories as well as regions. Moreover, the EVIS.AT platform manages access rights (e.g. which federal road authority is allowed to get which data for which region). Another important feature of the FCD platform is the periodic update of the road network graph ensuring that always the newest available graph is used for FCD processing. For managing different graph versions, we use the open source software Graphium[4].

Data privacy

Data privacy is one of the most critical issues when using FCD. In order to ensure the conformity of European as well as Austrian data protection regulations (on May, 25th 2018 the Austrian Data Protection Act - DPA[5] which is based on the European General Data Protection Regulation – GDPR[6] has entered into force), we have designed the FCD platform in a way that it processes anonymized data only. Therefore, we apply the following data protection measures:

- GNSS raw measurements are allowed to enter the platform as pre-anonymized data only, e.g. collected GNSS measurements may not contain any personal or vehicle identifiers, but only randomly assigned and periodically changing trip identifiers.
- After entering the FCD platform, all external, pre-anonymized identifiers are one-way re-hashed to random identifiers so that the platform never handles external identifiers.
- Fleet identifiers (if in the data) are removed on entry so that the data is managed as a general

[4] https://github.com/graphium-project/

[5] https://www.parlament.gv.at/PAKT/VHG/XXV/I/I_01761/fname_643605.pdf

[6] https://www.eugdpr.org/

data pool of vehicle trajectories without any possibility to identify trajectories by fleets, vehicles or persons. Trajectories are only attributed with generalized fleet characteristics (e.g., heavy goods vehicles or ambulance vehicles) for applying specific processing steps.

- Trajectories are matched solely against the higher-order road network (functional road classes 0-4) and GNSS measurements from minor roads are dismissed. This measure typically cuts start and end of GNSS trajectories so that vehicle start and destination addresses may not be reconstructed. A stand still of more than 3 minutes also cuts trajectories.
- Data sharing between federal road authorities is limited to anonymous link-based travel times and quality parameters. GNSS raw data never leaves the FCD platform.
- A contract between the fleet and the FCD platform operator specifies the agreed data usage rights. By contract, data usage is limited to traffic-related analyses; all other usages are excluded. The usage of derived travel time data is also limited to traffic-related purposes.

Two lawyers have independently checked the data protection measures and have confirmed that the applied approach for processing FCD adheres to the European GDPR as well as the Austrian DPA.

Quality measures

The overall objective of the Austrian national FCD platform aims at near real-time collection, processing and sharing of vehicle-originated FCD to be used for traffic information services of federal road authorities. Continuous monitoring of well-defined quality parameters is therefore a crucial pre-requisite for a successful deployment. Therefore, we introduced distinct quality measures being defined along the quality dimensions and objects proposed by the international standards ISO 21707 Intelligent transport systems – Integrated transport information, management and control – Data quality in ITS systems (ISO 21707, 2008) and ISO 19157 on Geographic Data Quality (ISO 19157, 2013). We propose the following quality measures along the quality dimensions *completeness*, *spatial accuracy* and *timeliness*. Table 1 gives an overview on the measures as well as their implementations.

Completeness

Completeness of a traffic service or data source can be measured in terms of physical coverage, e.g. measuring which parts of a road network are covered by the service. Firstly, in order to get a rough estimate of the amount of processed data per FCD source, we introduce the measure *successfullyProcessedGNSSMeasurements* as ‘share of all collected GNSS measurements being successfully processed by the FCD platform’. In order to gain knowledge about the physical coverage, we introduce the measure *successfullyMatchedKilometers* measuring the actual physical coverage with respect to different parts of the road network and different functional road classes. For a more detailed view, the parameter *FCDPenetrationRate* measures the share of FCD vehicles in relation to all vehicles for a distinct section of the road network (e.g. at a loop detector). In contrast to *successfullyMatchedKilometers*, this parameter allows for a detailed picture of the regional FCD coverage related to traffic volumes.

Spatial accuracy

For measuring the spatial quality of FCD, we use two measures. *averageSamplingInterval* measures the average time between two GNSS measurements being a relevant indicator for spatial accuracy of the trajectories. In addition, the *averageMatchingDistance* parameter defines how far the GNSS measurements on average are distant from the road network.

Timeliness

Since the data is used in near real-time systems, timeliness of data collection, transmission and processing is a key quality indicator. Therefore, we defined the following quality measures: *averageCollectionLatency* measures the average latency of the data when it enters the FCD platform (e.g. the timespan from the measurement of a GNSS location to the entry into the FCD platform) and *averageProcessingLatency* measures the average processing for the data from entering of raw data to sharing travel time values.

Table 1 – Implemented FCD quality measures

Dimension	Measure	Implementation/Evaluation
Completeness	*successfullyProcessedGNSS Measurements*	Measures the share of successfully processed GNSS measurements. Continuous evaluation.
Completeness	*successfullyMatchedKilometers*	Measures the coverage of the pre-defined road network. Continuous evaluation.
Completeness	*FCDPenetrationRate*	Measures the share of FCD vehicles compared to the total number of vehicles. Once a year evaluation for selected loop detectors.
Spatial accuracy	*averageSamplingInterval*	Measures the mean sampling interval between two GNSS measurements. Continuous evaluation.
Spatial accuracy	*averageMatchingDistance*	Measures the mean matching distance of GNSS measurements to the road network. Continuous evaluation.
Timeliness	*averageCollectionLatency*	Measures the latency of the collected FCD data on entering the FCD platform. Continuous evaluation.
Timeliness	*averageProcessingLatency*	Measures the latency of data processing in the FCD platform. Continuous evaluation.

Experiences from daily operation

Having its origins in the nationally funded Austrian flagship project FCD Testbed Salzburg, operation of the FCD platform has been started in 2013 with local fleets in the Federal State of Salzburg. Since then, the platform has been continuously extended, on the one hand regarding functionality and on the other hand regarding geographical coverage. Since 2016, the platform provides national FCD coverage and since 2017, the platform serves travel time data to other road authorities. In early 2018, the platform was ported from own hosting to the virtual cloud infrastructure of Open Telekom Cloud (OTC). Operation within a cloud infrastructure has the benefit of a completely virtualized environment, which is less error prone regarding hardware failures and allows easy scaling of computing and storage resources.

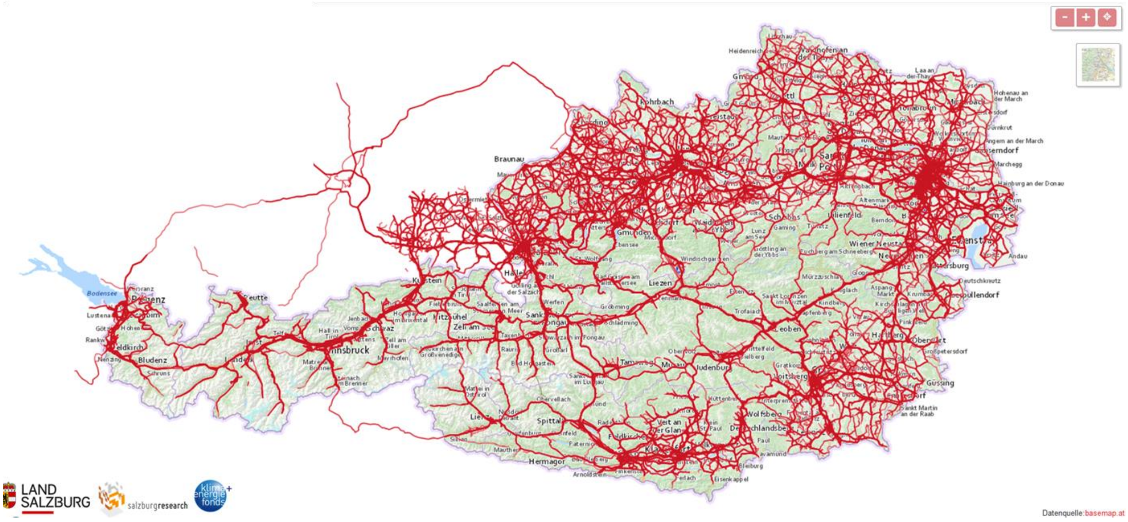

Figure 1: Visualization of the more than 80.000 daily vehicle trajectories being collected and processed

By the middle of 2018, the platform is connected to 13 different telematics service providers collecting FCD from dozens of different fleets. The fleet mix consists of logistic fleets, public transport fleets, taxis, service fleets, mobile care fleets, ambulance fleets and navigation devices of individual drivers. In terms of physical and temporal coverage of the road network, a well-chosen mix of fleets has been proven as successful since each fleet has its own specific characteristics and contributes to the overall physical coverage. On an ordinary weekday in May 2018, during peak hours (e.g. between 7 and 8 in the morning), more than 5.500 vehicles have been actively contributing FCD with the result of more than 80.000 daily trajectories consisting of 30 million GNSS measurements. This equals to 1.3 million matched kilometers on the road network (functional road classes 0 to 4, 46.116 km length) or 4.4 million directional travel times for road links (the covered road network consists of 233.718 mostly bi-directional road links). In order to determine physical coverage more reliably, we analyzed the FCD penetration rate for individual road links (we used links being equipped with a roadside loop detector). Presented FCD penetration rates in Figure 2 have been calculated as average from hourly FCD counts between 6 am and 8 pm in relation to average hourly traffic counts during different test weeks in 2016. The values show FCD penetration rates between 0.1 and 2.5 percent of the traffic volumes. Considering suggested penetration rates from literature (Breitenberger, Grüber, Neuherz, & Kates, 2004), the values give useful insights, on which roads the necessary penetration rates have already been reached and on which roads they have to be further improved. The presented figures from 2016 should be considered as first examples demonstrating the proposed quality measure *FCDPenetrationRate*. A more extensive evaluation is planned for autumn 2018, as the acquisition of additional FCD sources has been started early 2018 only and is still ongoing.

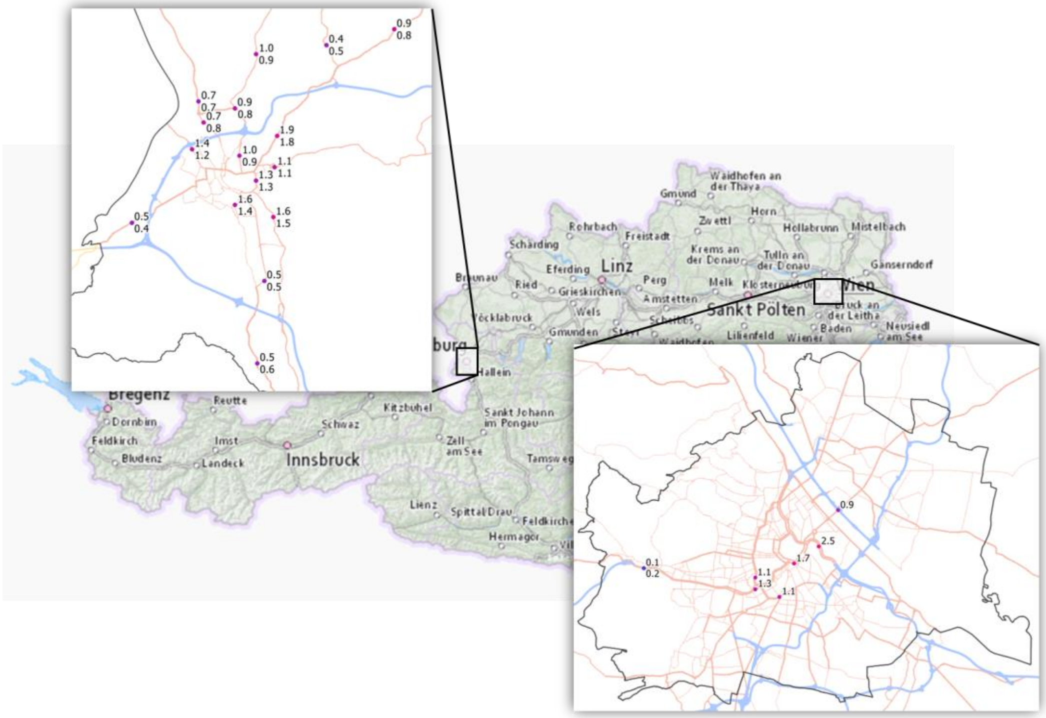

Figure 2: FCD penetration rates for selected road links in Salzburg and Vienna (values show percentages)

Another relevant quality measure are FCD sampling intervals in seconds. As shown in Figure 3, sampling intervals of current FCD sources are distributed between 1 and 80 seconds with peaks at 1 second (18 % of the data) and around 15 seconds. As clearly visible, a majority of the data is collected at sampling intervals less than 25 seconds guaranteeing good map matching quality (Rehrl et al., 2018). In the future, the parameter *averageSamplingInterval* will be calculated automatically per FCD source together with the other quality parameters *sucessfullyProcessedGNSSMeasurements*, *matchedKilometers*, *averageMatchingDistance*, *averageCollectionLatency* and *averageProcessingLatency* as part of a comprehensive quality monitoring system. This system will allow for continuous monitoring of the relevant operational parameters and will allow reacting upon critical deviations.

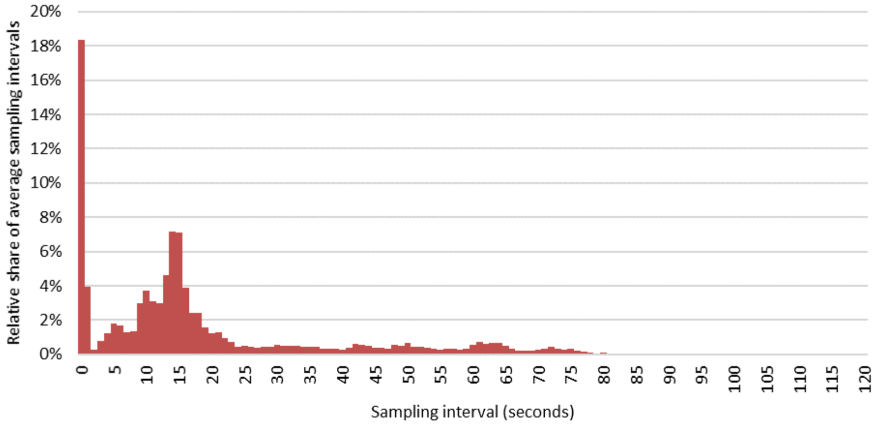

Figure 3: Distribution of sampling intervals over all vehicle fleets

Conclusions

This work describes the main concepts of the Austrian national FCD platform for collecting, processing and sharing floating car data. Establishing an own platform opens an alternative perspective on FCD for road authorities besides contracting FCD-based travel times from commercial traffic information service providers. The Austrian way of establishing a national FCD platform has proven successful in several ways. Firstly, the cooperative deployment and operation of a common FCD platform saves individual road authorities money and helps to assure common data quality standards. Secondly, the direct contracting of FCD raw data sources ensures comprehensive data usage rights including storage and publication of derived traffic information. Thirdly, road operators have direct control over data protection and data quality from the raw FCD sources to derived traffic information. This is especially useful when the traffic information serves as source for traffic planning or traffic management measures. Further plans include improvement of data coverage and data quality and turning the pilot operation into permanent operation.

Acknowledgements

This work has been funded by the Austrian Climate and Energy Funds as part of the EVIS.AT project.

References

1. Breitenberger, S., Grüber, B., Neuherz, M., & Kates, R. (2004). Traffic information potential and necessary penetration rates. *Traffic Engineering & Control*, *45*(11).
2. Brunauer, R., & Rehrl, K. (2014). Deriving driver-centric travel information by mining delay patterns from single GPS trajectories. In *Proceedings of the 7th ACM SIGSPATIAL Int. Workshop on Comp. Transp. Science - IWCTS '14* (pp. 25–30). New York, , USA: ACM Press.
3. ISO 19157. (2013). ISO 19157:2013 Geographic information - Data quality. ISO.
4. ISO 21707. (2008). Intelligent transport systems - Integrated transport information, management and control - Data quality in ITS systems. ISO - International Standards Organisation.
5. Rehrl, K., Brunauer, R., & Gröchenig, S. (2016). Collecting Floating Car Data with Smartphones: Results from a Field Trial in Austria. *Journal of Location Based Services*, *10*(1), 16–30.
6. Rehrl, K., Gröchenig, S., & Wimmer, M. (2018). Optimization and evaluation of a high-performance open-source map-matching implementation. In *Geospatial Technologies for All: Selected Papers of the 21st AGILE Conference* (pp. 251–270). Springer, Cham.
7. Thill, J.-C. (2000). Geographic information systems for transportation in perspective. *Transportation Research Part C: Emerging Technologies*, *8*(1), 3–12.
8. Viti, F., Hoogendoorn, S., Immers, L., Tampère, C., & Lanser, S. (2008). National Data Warehouse: How the Netherlands Is Creating a Reliable, Widespread, Accessible Data Bank for Traffic Information, Monitoring, and Road Network Control. *TRR: Journal of the TRB*, *2049*, 176–185.