Journal homepage: https://itce.vn.ua/en

Information Technologies and Computer Engineering | Vol. 23, No. 2. 2026

Article's History: Received: 09.01.2026 Revised: 08.05.2026
Accepted: 25.06.2026 Published: 03.08.2026

ISSN 1999-9941; e-ISSN 2078-6387
UDC 004.358
DOI: 10.31649/vitce/2.2026.90

Efficiency of the collection and analysis of vehicle telemetry data based on the modified European Data Format Plus standard

Ivan Bolhov

Postgraduate Student
National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute"
03056, 37 Beresteiskyi Ave., Kyiv, Ukraine
https://orcid.org/0009-0007-0710-5496

Yaroslav Klyatchenko*

PhD in Technical Sciences, Associate Professor
National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute"
03056, 37 Beresteiskyi Ave., Kyiv, Ukraine
https://orcid.org/0000-0003-4236-4059

Abstract. The relevance of this study is determined by the need to transition from closed proprietary formats to open containers capable of ensuring the integrity of multichannel time series during long-term testing. The aim of the work was to substantiate the use of the modified European Data Format Plus format for the collection and analysis of wheeled vehicle telemetry by developing a mechanism for the dynamic appending of information without disrupting the file structure. The article investigated the problem of implementing unified standards for the storage and exchange of telemetry information in automotive engineering. During the study, methods of comparative analysis of the technical specifications of binary formats and mathematical modelling of linear interpolation processes for converting digital values into physical units were applied. The results of the work demonstrated the high adaptability of the specified standard for recording vibration, temperature, and pressure parameters, which are typically characteristic of industrial diagnostic systems. An algorithmic scheme for updating the file header was developed and verified, allowing new data blocks to be added to existing structures subject to verification of the last timestamp and automatic adjustment of the number of records. This resolved the critical problem of information integrity in the event of sudden power failure or improper termination of the recording process. Experimental validation of the developed method during long-term monitoring sessions confirmed the preservation of high system performance and the absence of information loss during cyclic access to the storage medium. The approach proposed in the study also simplified further post-processing of data through the formation of a single integrated file for each testing cycle. Ultimately, this expanded the potential application of the investigated format in fields requiring reliable continuous collection of multiparametric telemetry. The practical value of the study lies in the use of the research findings during the transition to the concept of a single file per testing cycle, which increases the economic efficiency of data processing in mechanical engineering

Keywords: open data containers; binary data standards; automotive engineering; time-series recording; diagnostics of mobile systems; incremental information storage

Introduction

The trend towards the widespread use of intelligent transport systems, the introduction of autonomous driving protocols and the electrification of powertrains have led to an exponential increase in the volume of data generated by modern vehicles during testing. This trend necessitates a shift from fragmented proprietary formats and cumbersome text files, such as Comma-Separated Values, to a unified, open and energy-efficient standard for storing multi-channel time series. The general trend towards cost reduction in the automotive industry requires the implementation of binary containers capable of efficiently synchronising high-frequency signals from multiple sensors, whilst minimising the costs of storing and processing large volumes of data on cloud-based telemetry platforms.

Suggested Citation:
Bolhov, I., & Klyatchenko, Ya. (2026). Efficiency of the collection and analysis of vehicle telemetry data based on the modified European Data Format Plus standard. *Information Technologies and Computer Engineering*, 23(2), 90-99. doi: 10.31649/vitce/2.2026.90.

*Corresponding author (*k_yaroslav@ukr.net*)

Copyright © The Author(s). This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/)

The focus of automotive data standardisation since 2020 has shifted towards the creation of unified information exchange systems. F. Knapp & M. Simon (2023), in their review of standardisation processes in the automotive industry, emphasised that the lack of uniformity in formats is the main barrier to the development of intelligent mobility. The authors argued that the use of open and approved standards is not only a technical but also a strategic step towards ensuring compatibility between different component manufacturers. The practical implementation of Controller Area Network (CAN) bus data logging was examined by G. Soosaar & T. Lillerand (2025), who compared the efficiency of binary and text formats. The conclusions clearly indicated that binary structures allow for a significant increase in recording speed and a reduction in file size in wheeled vehicle telemetry. The importance of time synchronisation in distributed systems is highlighted in the work of R. Rane & P. Mulik (2025). It has been confirmed that the use of data formats with strictly fixed timestamps is essential for high-precision analysis of critical situations. This technical principle was first substantiated in the context of neurophysiological signal recording by B. Kemp *et al*. (1992) and is regarded as a fundamental advantage for ensuring the synchronisation of multi-channel data.

Modern signal processing methods for deep learning, analysed by G. Vishnu *et al*. (2023), require structured time series to improve the accuracy of predictive models. The authors emphasised that pre-processed binary data significantly accelerate the training phase of neural networks in tasks involving the modelling of wheeled vehicle behaviour. In the context of monitoring critical components, particularly lithium-ion batteries, P. Thakkar *et al.* (2024) proposed using multi-channel binary streams to detect faults in real time. Their work confirmed that the ability to embed event annotations directly into the data stream allows anomalies to be effectively recorded in parallel with the main current and voltage parameters. The scalability of such solutions in cloud environments was investigated by J. Scheuring & U. Bernhard (2020), who noted that open formats, which do not require specialised software licences, are the most promising for mass telemetry collection systems. Furthermore, the adoption of Regulation (EU) 2023/2854 (2023) establishes a legal framework obliging manufacturers to provide access to data in open formats, making the adaptation of existing standards, such as EDF+, a pressing scientific and practical task. Despite the existence of in-depth research in related fields, the use of the EDF (European Data Format) specifically in automotive engineering remains under-researched. As of 2025-2026, methods requiring attention include those for the direct transformation of CAN protocol identifiers into the EDF header structure directly on on-board computers. Existing approaches are predominantly focused on subsequent data conversion, which prevents the utilisation of the format's advantages for online diagnostics and data transmission over channels with limited bandwidth. There is also a need to develop header compression algorithms for cases where the number of signals runs into the thousands, which is characteristic of modern autonomous vehicle control systems.

The aim of the study was the theoretical substantiation and development of a methodology for adapting the EDF format for the collection of test information from onboard systems of wheeled vehicles. To achieve this aim, it was necessary to solve the following objectives: to develop an algorithmic model for transforming messages into a multi-channel EDF structure, taking into account different sampling frequencies, and to design a mechanism for processing asynchronous events in a vehicle network using annotations. The algorithmic model for message transformation is based on determining the base time period dt and calculating the sampling rate for each channel. For example, when synchronising low-frequency and high-frequency (1,000 Hz) signals, the model forms a structure where 1 sample of the low-frequency signal corresponds to 100 samples of the high-frequency signal within each data block. This ensures strict phase synchronisation of the parameters without the need for resource-intensive resampling. The mechanism for processing asynchronous events implements an "event-record-link" scheme through the use of a dedicated annotation channel. Upon event activation, the algorithm generates a tag and inserts it into the current data block with a precise timestamp corresponding to the physical signal. This approach allows critical vehicle operating modes to be automatically indexed directly within the binary telemetry array. Completing these tasks will enable the creation of a unified data collection toolkit, which will help improve diagnostic accuracy and accelerate the processing of large telemetry data sets.

Materials and Methods

The methodological basis of the study was a comprehensive approach to the collection and structuring of diagnostic data obtained during vehicle testing. To ensure the results were easily reproducible, the experimental part was based on the use of a data logging software programme responsible for creating, recording and verifying the integrity of files. This logger represents a specialised software package designed for cross-platform emulation of data collection processes, utilising low-level direct file system access mechanisms to precisely control the structure of the stored arrays. It contains a built-in ring buffer module, which allows data packets to be temporarily buffered before being written to the physical storage medium, minimising the load on the input-output subsystem and preventing data loss during peak loads.

The recording of parameters was carried out using data streams with defined characteristics, such as the name of the physical quantity and the sampling rate, which varied within the experiment according to the dynamics of the processes under investigation. This allowed data to be recorded and processed in both real-time and pseudo-real-time modes, where the load on the onboard CAN bus is simulated. Thus, file processing is decoupled from the direct data source but allows the simulation of a data stream fully corresponding to real-world tests of wheeled vehicles under road or laboratory conditions. The EDF+ (European Data Format Plus) format was selected as the central element of the data storage architecture; this is regulated by official specifications and normative documents in the field of standardisation of medical and general technical time series and is established as the basic protocol for the exchange of biomedical and technical time series under the European Data Format (2003), which is supported by the international scientific community to ensure the compatibility of data recording systems. The use of this standard is due to its strictly deterministic structure and its ability to store multi-channel parametric data together with events and annotations in a single container. The file generation process requires the use of a 16-bit signed digital representation of signals, which provides the necessary resolution within the established dynamic range.

During the mathematical conversion of physical quantities into digital values, a linear scaling method was applied, which strictly complies with the standard's requirements regarding the mandatory preservation of transformation coefficients, as presented by B. Kemp & J. Olivan (2003). Particular attention is paid in the study to the American Standard Code for Information Interchange (ASCII) header configuration, which records the technical characteristics of each channel, including physical dimensions, types of analogue and digital filtering, as well as signal limit values. The experimental strategy involves cyclic testing under specified operating conditions, depending on the number of data transmission channels and the number of records in a single packet. For the mathematical interpretation and scaling of measurement signals, the study employs a linear interpolation method, which allows 16-bit integers read from a file to be transformed into actual physical quantities in accordance with the standard specification. The conversion formula is as follows:

$$P = \frac{P_{max} - P_{min}}{D_{max} - D_{min}} \times (D - D_{min}) + P_{min}, \qquad (1)$$

where P – the physical value, D – the digital value from the file, P_{max} and P_{min} – the physical limits specified in the header, D_{max} and D_{min} – the digital limits specified in the header.

In addition, for memory resource planning during long-term experiments, a method of deterministic calculation of the output file size is used according to the following formula:

$$S_{total} = 256 + (ns \times 256) + (N \times \sum_{i=1}^{ns} S_i \times 2), \qquad (2)$$

where 256 – the size of the static header, ns – the number of channels, N – the number of recording blocks, S_i – the number of samples of the i-th channel within a block.

Summing the sampling frequencies across the entire set of channels allows the frequency differences between channels to be taken into account. The final file size is the product of the experiment's duration and the total bandwidth of all measurement channels. To ensure data integrity during long sessions, a proprietary algorithm is used to iteratively append data to an already open and partially filled file. This procedure is implemented by the software logger by dynamically updating the number of records field in the header after the completion of each individual test stage. This approach avoids file system fragmentation, eliminates the need for constantly reopening file descriptors, and ensures the formation of a single structured array for the entire test session, which fully complies with the fundamental principles of reliable data archiving in engineering systems.

Results and Discussion

The issue of data standardisation has gained importance in the field of monitoring dynamic vehicle parameters. Current research in automotive engineering confirms that the effectiveness of computer modelling of complex assemblies, particularly tyres, directly depends on the architecture of the data collection and storage system. The study found that modern vehicles require the integration of numerous sensors and transducers, which together generate a large volume of high-density telemetry data. An analysis of the effectiveness of tyre modelling tools based on the use of embedded telemetry controllers showed that integrating sensors directly into the wheel structure generates massive streams of multi-channel signals: pressure, temperature and deformation forces. In this context, it is critically important to resolve the issue of precise data synchronisation from such heterogeneous sources as pressure, temperature, vibration and deformation force sensors, as any temporal desynchronisation between them distorts the results of the mathematical analysis of motion dynamics.

An analysis of the performance of platforms for processing large spatial arrays, conducted by F. Hu *et al*. (2018), found that most systems require resource-intensive pre-processing of data, which highlights the importance of selecting an optimal storage format, such as EDF+, to minimise loading costs. Thus, whilst traditional formats (e.g., Measurement Data Format) are tailored to the specific needs of industrial diagnostic equipment, the implementation of EDF+ in bus telemetry analysis processes can ensure a higher rate of information exchange between research groups. The cross-platform nature of the format, as argued by B. Kemp & J. Olivan (2003), becomes a critical factor for implementing a monitoring system, as it allows data from embedded telemetry controllers to be seamlessly integrated into cloud environments for further mathematical analysis and refinement of tyre wear models.

The European Data Format (EDF) and its enhanced version, EDF+, are widely used formats in the field of medical technology for the storage and exchange of biosignals. The results of EEG (electroencephalogram) or polysomnography are generated specifically as EDF files. The classic EDF was developed by B. Kemp *et al.* (1992) with the aim of introducing a single standard for recording and processing information to replace the proprietary formats of various medical equipment manufacturers (Smrity, 2025). Although the EDF format was developed specifically for medical purposes, its structure is a universal container for multi-channel time series. The format has found application in several related fields that require the synchronous recording of data from multiple sensors. In fact, any system requiring the recording of physical parameters over time can use EDF. The format is widely used in engineering to monitor the condition of structures and the operation of equipment, as it allows for the synchronous recording of vibrations, temperature and pressure within a single file. In seismology and oceanography, this format is convenient for data exchange between stations, as all units of measurement are specified directly in the header of each channel. In professional sport, EDF is used to analyse athletes' movements alongside heart rate data, which helps assess fatigue and the effectiveness of training sessions. EDF is the standard for lie detectors, as it easily combines data on breathing, skin activity and eye movements. The format's popularity stems from its compactness compared to text tables and the fact that all sensor configuration information is stored directly within the file, preventing errors during analysis.

The automotive industry faces similar challenges, which previously led to the adoption of a unified format for data storage and processing. The need for standardisation stems from the trend towards simplifying information exchange processes between manufacturers of different components, which leads to the acceleration of research and production processes in the industry. The scientific debate regarding the effectiveness of EDF+ in non-medical sectors, particularly in automotive engineering and the monitoring of complex technical systems, highlights both the format's advantages and its critical limitations compared to alternative standards. Using the EDF structure allows telemetry data from embedded controllers to be consolidated into a single file with a clear description of the units of measurement and sampling rate for each channel. This resolves the issue of compatibility between on-board data acquisition systems and standalone software suites for modelling, thereby enhancing the overall efficiency of computational tools. Research by A. Schlögl & C. Brunner (2008), dedicated to the development of the BioSig library, indicates that the simplicity of the EDF structure is its main advantage for long-term archiving; however, the author emphasises the significant advantage of the GDF (General Data Format) in terms of supporting 32-bit data and flexible event description. Compared to the MDF (Measurement Data Format), which is the standard in the automotive industry, EDF+ demonstrates less flexibility when working with dynamic metadata, yet gains significantly in cross-platform compatibility thanks to its ASCII text header, as analysed in detail in the works of B. Kemp & J. Olivan (2003) when justifying the transition from the classic EDF to the EDF+ modification.

The structure of an EDF file is strictly deterministic. The entire file consists exclusively of ASCII characters (header) and 16-bit integers (data). This structure makes the format platform-independent. The header is a text block of ASCII characters that can be read without the use of specialised software using any text editor. The header is divided into a fixed 256-byte section, containing the format version, document name, start date of the recording, number of signals, etc., and a variable section, the size of which depends on the number of signals and includes the technical specifications of each channel: signal name, sensor type, units of measurement, physical minimum/maximum values, digital minimum/maximum values, and filter parameters. The structure of the fixed part of the EDF file header is shown in Table 1.

Table 1. Distribution of bytes in the EDF file header structure fixed part of the header

Fixed part of the header			
Offset (bytes)	**Length (bytes)**	**Content**	**Description**
0	8	version	Format version (for EDF always "0")
8	80	patient_id	Local file identifier
88	80	record_id	Local recording identifier
168	8	startdate	Start date (dd.mm.yy)
176	8	starttime	Start time (hh.mm.ss)
184	8	header_bytes	Number of bytes in the header (including the dynamic part)
192	44	reserved	Reserved (for EDF+, the discontinuity type is specified here)
236	8	num_records	Number of data records, −1 if unknown
244	8	duration	Duration of one record in seconds
252	4	num_signals	Number of signals (channels) in the file
Variable part of the header for each channel			
256	16	Channel names	Signal types (for example, EEG Fp1 or ECG)
16	80	Sensor type	Description of electrodes or transducers
80	8	Physical unit	Measurement units (uV, Ohm, degree, etc.)
8	8	Physical minimum	Minimum signal value in physical units
8	8	Physical maximum	Maximum signal value in physical units
8	8	Digital minimum	Minimum value after ADC conversion (usually −32768)
8	8	Digital maximum	Maximum value after ADC conversion (usually 32767)
8	80	Filtering	Description of filter settings (HPF, LPF, Notch)

Table 1. Continued

Fixed part of the header			
Offset (bytes)	**Length (bytes)**	**Content**	**Description**
80	8	Number of samples	Number of values for each signal in one block
8	32	Reserved	Additional field for each channel

Source: B. Kemp *et al*. (1992)

The data array is placed directly after the file header in the form of an ordered sequence of binary blocks containing digital signal values. A data block is the principal fragment of the file containing digital signals. Unlike conventional tables, where data are recorded row by row, in EDF they are grouped into segments. The structure of each block provides for the sequential grouping of blocks by each channel within a defined time interval. This recording method is known as block multiplexing. The classical EDF format has certain limitations caused by the strict static nature of its structure, where all recording channels must have identical duration within a single data block, making efficient recording of signals with different temporal characteristics impossible. In contrast, the EDF+ specification, introduced by B. Kemp & J. Olivan (2003), implements the concept of discontinuous recordings and variable sampling frequencies for different signal types, allowing both high-frequency and low-frequency signals to be integrated into a single file.

A fundamental distinction lies in the annotation mechanism: whereas in the original EDF format event annotations required the creation of separate text protocols, EDF+ implements specialised annotation channels that allow timestamps and textual comments to be stored directly within the binary stream. This ensures precise synchronisation of simultaneous events, such as technical malfunctions, with the corresponding signal segments. In addition, EDF+ expands the informational capacity of the header by introducing standardised fields for file identification and technical characteristics of the equipment. Despite its more complex internal organisation, EDF+ retains structural similarity to the classical EDF format, thereby ensuring backwards compatibility for reading EDF+ files with EDF editors, although specific extensions such as annotations are ignored. Thus, the transition to EDF+ transformed the format from a simple biosignal container into a comprehensive system for storing multiformat data suitable for automated analysis and long-term archiving. A comparison of the characteristics of the EDF and EDF+ formats is presented in Table 2.

Table 2. Comparison of EDF and EDF+ formats

Characteristic	EDF	EDF+
Data type	16-bit Integer	16-bit Integer
Text encoding	ASCII	ASCII
Annotations	Absent	Supported (annotation channel)
Compatibility	Basic	Backwards compatibility with EDF
Field of application	EEG, ECG, polysomnography	Any time series data

Source: B. Kemp & J. Olivan (2003)

The practical implementation of the mathematical framework in the software logger confirmed the stability of the linear data transformation. By using integer arithmetic at the recording stage, it was possible to maintain a high speed of data aggregation, whilst conversion to physical quantities using formula (1) was performed exclusively during post-processing or data visualisation. The practical calculation of the file size using formula (2) allowed to predict disk space usage and respond in a timely manner to the threat of data loss in the event of insufficient storage capacity. The use of integer arithmetic ensured a recording speed of 2.4 Mbit/s for a stream from 32 analogue channels. The use of linear transformation according to formula (1) ensured the conversion of digital values (in the range from -32768 to 32767) into physical units with an error not exceeding 0.01% of the sensor's full scale. For example, for a pressure sensor with a range of 0-100 kPa, the quantisation step was 1.5 Pa, which fully meets the accuracy requirements for brake system diagnostics. Predicting disk space usage using formula (2) revealed that one hour of continuous telemetry recording (at a total sampling rate of 5,000 Hz) occupies 34.3 MB of memory. This enables autonomous testing for up to 230 hours on a standard 8 GB storage device, which is critical for endurance testing of vehicles. Furthermore, the test results highlighted the key advantages of EDF+ for automotive engineering, including high-speed processing of large data sets through block multiplexing, cross-platform compatibility of the text header, and the absence of a need for proprietary libraries. It was also demonstrated that this approach to telemetry data collection can be effectively applied to improve the efficiency of vehicle performance analysis by continuously recording real-world operational loads, thereby significantly enhancing algorithms for predicting the remaining service life of components and optimising maintenance cycles.

Despite its long-standing dominance in the field of medical diagnostics, the architecture of the EDF and EDF+ formats contains a number of fundamental shortcomings stemming from outdated technical standards. The main limiting factor is the 16-bit nature of data storage, which restricts the signal's dynamic range to just 65,536 levels. For modern high-precision equipment with 24-bit or 32-bit resolution, this means an irreversible loss of detail during conversion. Furthermore, the format's mathematical model supports only a linear transformation between the digital code and the physical quantity, making it impossible to directly store data from sensors with logarithmic or other non-linear characteristics. The lack of built-in compression algorithms leads to excessive use of disk space, as data is stored in its original "raw" form without volume optimisation. The MDF format solves this problem by introducing built-in time-series archiving, but in turn creates the problem of asynchronous reading of different data channels (Marten *et al.*, 2019). Compression makes it possible to reduce file size, which is efficient for long-term storage, but requires synchronisation during analysis. The fixed length of textual fields prevents the inclusion of detailed metadata, while support exclusively for ASCII characters restricts the use of national alphabets in service information.

A particular issue is the specific method of integrating annotations directly into the body of the file between measurement blocks. This structure makes any editing of labels or correction of errors in the record a complex operation, requiring the entire data set to be completely rewritten rather than allowing for local adjustments to values. Research by R. Huttunen *et al.* (2023) indicates that although polysomnography data can be exported to open formats such as EDF, there are limitations regarding the verification of data blocks and readability. However, when processing large volumes of data during the analysis of wheeled vehicle behaviour, such limitations are offset by the absence of a need to read data directly during testing. These limitations arise from different causes but are united by the need to make changes to the file structure, as well as the requirement for specialised software for both recording and reading information. These features of the EDF format present specific challenges for developers, as any deviation from the 1992 standard renders the file incompatible with most viewers. This forces engineers to seek a compromise: either strictly adhere to the restrictions or create their own format extensions that can only be read by specific software tools. However, there are limitations that do not require changes to the file structure. One of the key challenges in analysing telemetry data is synchronising numerical indicators (CAN bus, GPS) with the video stream. Unlike EDF/EDF+ files, video files do not usually contain embedded metadata regarding the exact date and time of the start of recording in their internal headers. To address this issue within the modified EDF+ standard, it was proposed to use a unified timestamp encoding directly in the video file name. For example, a file named "UA_VEHICLE01_25-MAR-2026_12h30m05.500s_Cam1.mp4" indicates that the vehicle video recording began on 25 March 2026 at 12:30:05 and 500 milliseconds. This approach allows specialised software to automatically read the start time from the video filename and the contents of the EDF file, ensuring millisecond accuracy when playing back and analysing critical events. The main advantage of this innovation is full compatibility with recording and playback devices developed previously.

One of the most critical drawbacks of the format is the inability to dynamically append data to the end of an existing file in real time without the risk of corruption if the file has already been closed. This problem stems from several technical reasons. Firstly, it is the dependency on the final value of the number of records per block, which is set in the header. The problem is that the viewer programme first reads this value and only then begins to display the data. If a failure occurs during recording, such as a power cut, and the programme does not have time to update this number in the header, the file will be considered empty or corrupted, even if gigabytes of data are physically present on the disk. Secondly, the EDF records data not by individual channels, but in time blocks. A single block contains all samples from all channels for one second. To add a new block, the programme must know the exact physical offset from the start of the file. Any one-byte error during writing leads to a shift in the entire data grid, after which the programme will start interpreting random bytes as signals. Since EDF requires structural integrity, developers are often forced to create temporary buffers, which increases the risk of losing the last few minutes of recording if the software shuts down incorrectly. Implementing the write-back mechanism in EDF+ format log files ensures compliance with the "one file – one experimental session" principle. This approach eliminates the need for subsequent concatenation of fragmented data, which significantly improves the reliability of export procedures and minimises the likelihood of artefacts or errors occurring during digital post-processing of the results.

The issue of data integrity during write operations has also been examined in the context of real-time system development. For example, research led by N. Rishikesh (2022) on vibration monitoring highlights that formats with a fixed header structure require specific caching algorithms to prevent file corruption during sudden power failures. The mechanism proposed in this work for verifying the latest timestamp and dynamically updating the number of records correlates with approaches used in high-performance loggers; however, adapting these methods specifically for EDF+ allows for full compatibility with existing diagnostic software without resorting to complex hierarchical structures.

In the context of EDF+, simply appending bytes to the end of the file will render it unreadable for standard viewers, as the header will state that the file contains, for example, 100 records, although in fact there are already 200. Correcting the header is critical for compliance with the standard. The process of updating a record in an already created and closed EDF+ file requires a specialised algorithmic approach due to the structure of this format's header. Appending after reopening the file requires verification of the last timestamp of the last data block. This prevents time gaps or overlaps, which is critical for correct analysis. The overwrite mode without file reconstruction prevents the loss of previously accumulated data in the event of a sudden power failure or a critical termination of the recording process during the current session. When implementing append mode in EDF+, it should be noted that the format has a fixed structure; thus, it is important that the sampling parameters and number of channels in the new data fully match those declared when the file was originally created. As the EDF+ specification requires the number of data records to be clearly specified in the header, when opening the file for appending, the current header must be read in order to update the "number of data records" field after recording is complete. If 100 new one-second data blocks have been added to the existing 500, the value of "number of data records" should be updated to 600, left-aligned and padded with ASCII space characters. The process of adding new data blocks is shown in Figure 1.

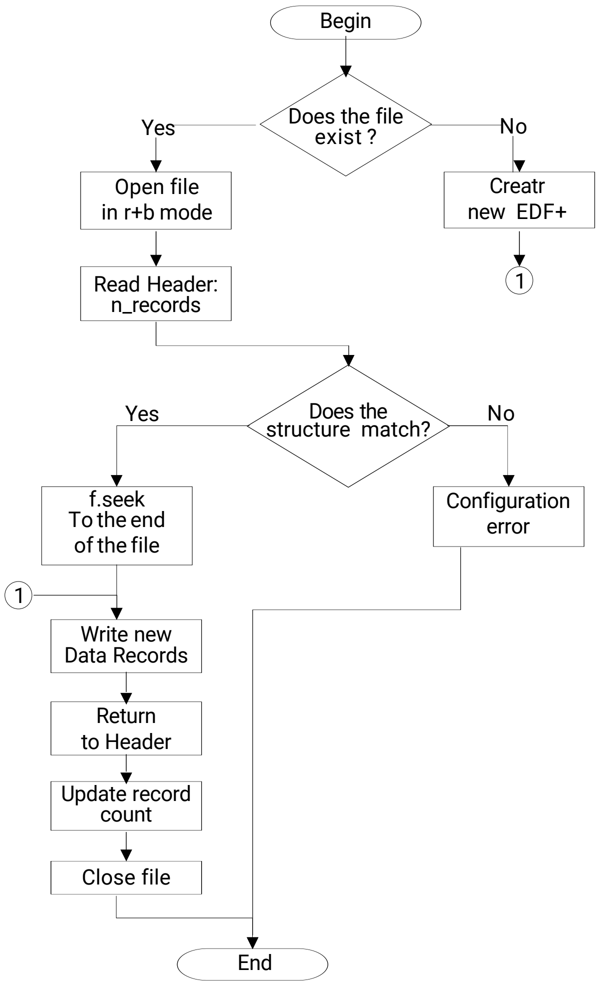

Figure 1. Algorithmic scheme for appending data to an EDF+ file

Source: developed by the authors

The issue of scalability and the processing of large datasets in EDF format is becoming particularly pressing in the context of modern cloud computing. Research by S.S. Sahoo *et al.* (2014) highlights that working with "big data" in clinical research presents specific challenges related to the distributed analysis of electrophysiological signals. The authors note that the architectural rigidity of EDF can cause computational delays when attempting to process files in parallel in cloud environments, which reinforces the argument for the need to improve data access methods, in particular by optimising header structures for rapid indexing. The work by J.R. Chandiramani *et al.* (2014) emphasises that telemetry acquisition systems are an integral part of the work for any racing team or vehicle manufacturer. The authors note that telemetry is crucial during the development phase, as it allows designs to be modified and parameters adjusted to improve performance and efficiency. The SB-RNLDC (Subsampled and Balanced Recurrent Neural Lossless Data Compression) method, proposed by P. Ramalingam *et al.* (2022), addresses the problem of limited bandwidth and storage capacity for telemetry data by implementing a deep learning architecture that combines signal pre-centring with lossless coding based on probability balancing. However, this method requires significant computational power and cannot guarantee the fidelity of the stored information. G. Wojnar *et al.* (2021) shift the focus to the engineering domain, examining the interpretation of vibration signals for structural health monitoring. The authors demonstrated that the use of synchronised time series from multiple sensors allows the identification of structural defects at an early stage, even in the presence of operational noise.

The accuracy of the diagnosis depends directly on the phase synchronisation of the signals, which is technically ensured by the block-based recording structure characteristic of EDF. As demonstrated by A.E. Oskouei *et al.* (2025), an approach based on the analysis of symmetric time patterns and the detection of their asymmetric deviations allows a neural network to be used to identify erroneous signal segments. Nevertheless, testing of wheeled vehicles is not limited to static data of a single type, and such filtering cannot guarantee that the detected anomaly is indeed a signal processing error rather than a change in the signal itself. A comparison of these results confirms the theory regarding the high adaptability of the EDF+ container for industrial monitoring tasks. The analysis shows that whilst specialised engineering formats often require the use of paid proprietary libraries, the implementation of safe overwriting methods in EDF+ makes this format competitive for synchronous data collection in field test conditions. In the work of N. Inoue & T. Konishi (2018) it is emphasised that validation of algorithms for autonomous driving systems is an extremely important yet highly labour-intensive process. However, the authors highlight a critical issue: due to the complexity of autonomous driving systems, the data used for modelling must be of high quality and quantity. The use of the EDF+ format makes it possible to satisfy these requirements owing to its capability to store massive streams of multichannel signals with temporal synchronisation.

Thus, resolving the issue of correctly modifying the header after the file is closed helps to mitigate one of the main architectural shortcomings of the standard identified by the researchers, whilst ensuring the reliability of data storage in line with modern industrial standards. The proposed algorithm ensures strict compliance with the ISO 60601-1 specification (International Organization for Standardization, 2005), as it performs atomic updates of the data record counter and maintains the consistency of the EDF+ annotation channel, thereby preventing desynchronisation during long-term monitoring.

Conclusions

The conducted study confirms that the European Data Format and its EDF+ modification, despite their medical origin, have evolved into a universal tool for recording multichannel time series in complex technical systems. Analysis of the format structure demonstrates that the combination of a textual ASCII header and a binary array of 16-bit data ensures a high level of cross-platform compatibility, which is critically important for the unification of data exchange between component manufacturers in the automotive industry and industrial monitoring. The application of the mathematical model of linear interpolation allows digital values to be efficiently converted into physical quantities; however, the rigid header architecture and the limited range of 16-bit encoding remain the principal limiting factors when working with high-precision, high-resolution sensors. Within the scope of the conducted research, particular attention was devoted to solving the problem of data integrity during the dynamic appending of blocks to existing EDF+ files. The application of the developed algorithmic approach, based on verification of the last timestamp and correct updating of the record count field in the header, ensured reliable logging according to the principle of "one session – one file". During the experimental validation, it was demonstrated that this method completely eliminates the risks of file structure corruption during sudden power failures and removes the need for subsequent complex concatenation of data fragments. This made it possible to position the EDF+ standard as a competitive open solution compared with proprietary engineering formats, particularly MDF. Furthermore, the developed algorithm confirmed its stability during continuous recording of high-frequency multichannel time series without reducing the overall system performance.

Analysis of the obtained vehicle telemetry data confirmed the effectiveness of the EDF+ format for synchronising characteristics with different frequencies, such as low-frequency engine temperature fluctuations and high-frequency suspension vibrations, within a single time scale. The implementation of an onboard system event annotation mechanism enabled the automatic identification of moments when electronic systems were activated together with the corresponding changes in physical telemetry parameters, thereby providing a comprehensive approach to assessing the technical condition of a vehicle under real operating conditions. The prospect of further research into the integration of EDF+ logging into a specialised monitoring and telemetry system for wheeled vehicles will enable the creation of a convenient data exchange interface while complying with standardised requirements.

Acknowledgements

None.

Funding

None.

Conflict of Interest

None.

References

[1] Chandiramani, J.R., Bhandari, S., & Hariprasad, S.A. (2014). Vehicle data acquisition and telemetry. In *Proceedings of the 2014 fifth international conference on signal and image processing* (pp. 187-191). Bangalore: IEEE. doi: 10.1109/ICSIP.2014.35.

[2] European Data Format. (2003). *EDF+ specification*. Retrieved from https://www.edfplus.info/specs/edfplus.html.

[3] Hu, F., Xu, M., Yang, J., Liang, Y., Cui, K., Little, M.M., Lynnes, C.S., Duffy, D.Q., & Yang, C. (2018). Evaluating the open-source data containers for handling big geospatial raster data. *ISPRS International Journal of Geo-Information*, 7(4), article number 144. doi: 10.3390/ijgi7040144.

[4] Huttunen, R., Rusanen, M., Nikkonen, S., Korkalainen, H., & Kainulainen, S. (2023). A harmonized and interoperable format for storing and processing polysomnography data. *arXiv*. doi: 10.48550/arXiv.2402.06702.

[5] Inoue, N., & Konishi, T. (2018). Data recording technology for autonomous driving/advanced driver assistance. *DENSO TEN Technical Review,* 3, 3-9.

[6] International Organization for Standardization. (2005). *IEC 60601-1:2005*. Retrieved from https://webstore.iec.ch/en/publication/2606.

[7] Kemp, B., & Olivan, J. (2003). European data format "plus" (EDF+), an EDF alike standard format for the exchange of physiological data. *Clinical Neurophysiology*, 114(9), 1755-1761. doi: 10.1016/s1388-2457(03)00123-8.

[8] Kemp, B., Värri, A., Rosa, C.A., Nielsen, K.D., & Gade, J. (1992). A simple format for exchange of digitized polygraphic recordings. *Electroencephalography and Clinical Neurophysiology*, 82(5), 391-393. doi: 10.1016/0013-4694(92)90009-7.

[9] Knapp, F., & Simon, M. (2023). Standardization of project management practices of automotive industry suppliers – systematic literature review. *Technical Bulletin*, 17(3), 432-439. doi: 10.31803/tg-20230504094426.

[10] Marten, D., Meyer, H., & Heuer, A. (2019). *Database support for automotive analysis*. Retrieved from https://eprints.dbis.informatik.uni-rostock.de.
[11] Oskouei, A.E., Kaveh, M., Hernando-Gallego, F., & Martín, D. (2025). Hybrid time series transformer – deep belief network for robust anomaly detection in mobile communication networks. *Symmetry*, 17(11), article number 1800. doi: 10.3390/sym17111800.
[12] Ramalingam, P., Mehbodniya, A., Webber, J.L., Shabaz, M., & Gopalakrishnan, L. (2022). Telemetry data compression algorithm using balanced recurrent neural network and deep learning. *Computational Intelligence and Neuroscience*, 2022(1), article number 4886586. doi: 10.1155/2022/4886586.
[13] Rane, R., & Mulik, P. (2025). Leveraging open table formats for scalable and flexible data infrastructure. *International Journal of Science and Research (IJSR)*, 14(6), 1219-1220. doi: 10.21275/sr25616222400.
[14] Regulation (EU) 2023/2854 of the European Parliament and of the Council "On Harmonised Rules on Fair Access to and use of data and Amending Regulation (EU) 2017/2394 and Directive (EU) 2020/1828 (Data Act)". (2023, December). Retrieved from http://data.europa.eu/eli/reg/2023/2854/oj.
[15] Rishikesh, N. (2022). Vibration monitoring system in industrial machinery using fast Fourier transform. *International Journal of Science and Research (IJSR)*, 11(5), 1738-1741. doi: 10.21275/sr22523164451.
[16] Sahoo, S.S., Jayapandian, C.P., Garg, G., Kaffashi, F., Chung, S., Bozorgi, A., Chen, C.-H., Loparo, K., Lhatoo, S.D., & Zhang, G.-Q. (2014). Heart beats in the cloud: Distributed analysis of electrophysiological "Big Data" using cloud computing for epilepsy clinical research. *Journal of the American Medical Informatics Association*, 21(2), 263-271. doi: 10.1136/amiajnl-2013-002156.
[17] Scheuring, J., & Bernhard, U. (2020). Time-sensitive networking in the automotive industry. *ATZ Electronics Worldwide*, 15, 54-58. doi: 10.1007/s38314-020-0236-9.
[18] Schlögl, A., & Brunner, C. (2008). BioSig: A free and open-source software library for BCI research. *Computer*, 41(10), 44-50. doi: 10.1109/mc.2008.407.
[19] Smrity, M.I. (2025). *Challenges and methods in biomedical signal quality evaluation: A multimodal approach*. (Master's thesis, Tampere University, Tampere, Finland).
[20] Soosaar, G., & Lillerand, T. (2025). An energy-efficient battery monitoring and logging system for agricultural robotics with CAN bus integration. *Environmental and Climate Technologies*, 29(1), 156-170. doi: 10.2478/rtuect-2025-0011.
[21] Thakkar, P., Khatri, S., Dobariya, D., Patel, D., Dey, B., & Singh, A.K. (2024). Advances in materials and machine learning techniques for energy storage devices: A comprehensive review. *Journal of Energy Storage*, 81, article number 110452. doi: 10.1016/j.est.2024.110452.
[22] Vishnu, G., Kaliyaperumal, D., Pati, P.B., Karthick, A., Subbanna, N., & Ghosh, A. (2023). Short-term forecasting of electric vehicle load using time series, machine learning, and deep learning techniques. *World Electric Vehicle Journal*, 14(9), article number 266. doi: 10.3390/wevj14090266.
[23] Wojnar, G., Burdzik, R., Wieczorek, A.N., & Konieczny, Ł. (2021). Multidimensional data interpretation of vibration signals registered in different locations for system condition monitoring of a three-stage gear transmission operating under difficult conditions. *Sensors*, 21(23), article number 7808. doi: 10.3390/s21237808.

Ефективність збору та аналізу телеметричних даних транспортних засобів на базі модифікованого стандарту European Data Format Plus

Іван Болгов

Аспірант
Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського»
03056, просп. Берестейський, 37, м. Київ, Україна
https://orcid.org/0009-0007-0710-5496

Ярослав Клятченко

Кандидат технічних наук, доцент
Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського»
03056, просп. Берестейський, 37, м. Київ, Україна
https://orcid.org/0000-0003-4236-4059

Анотація. Актуальність роботи зумовлена потребою переходу від закритих пропрієтарних форматів до відкритих контейнерів, здатних забезпечити цілісність багатоканальних часових рядів під час тривалих випробувань. Метою роботи було обґрунтування використання модифікованого European Data Format Plus формату для збору та аналізу телеметрії колісної техніки шляхом розробки механізму динамічного дозапису інформації без порушення структури файлу. У статті було досліджено проблему впровадження уніфікованих стандартів зберігання та обміну телеметричною інформацією в автомобільній інженерії. Під час дослідження було застосовано методи порівняльного аналізу технічних специфікацій бінарних форматів, математичне моделювання процесів лінійної інтерполяції для перерахунку цифрових значень у фізичні одиниці. Результати роботи продемонстрували високу адаптивність зазначеного стандарту для фіксації параметрів вібрації, температури та тиску, що зазвичай притаманне промисловим діагностичним системам. Розроблено та верифіковано алгоритмічну схему оновлення заголовка файлу, яка дозволяє додавати нові блоки даних у вже існуючі структури за умови перевірки останньої часової мітки та автоматичного корегування кількості записів. Це вирішило критичну проблему цілісності інформації при раптовому вимкненні живлення або некоректному завершенні процесу реєстрації. Експериментальна апробація розробленого методу під час тривалих сесій моніторингу підтвердила збереження високої швидкодії системи та відсутність втрат інформації при циклічному зверненні до накопичувача. Запропонований у дослідженні підхід також дозволив спростити подальшу постобробку даних завдяки формуванню єдиного цілісного файлу для кожного випробувального циклу. У підсумку це розширило потенціал застосування досліджуваного формату в галузях, де вимагається надійний безперервний збір багатопараметричної телеметрії. Практична цінність дослідження полягає у використанні даних дослідження при переході до концепції одного файлу на випробувальний цикл, що підвищує економічну ефективність обробки даних у машинобудуванні

Ключові слова: відкриті контейнери даних; бінарні стандарти даних; автомобільна інженерія; фіксація часових рядів; діагностика рухомих систем; інкрементне збереження інформації