

# Management von großen Sensordatenmengen für die Digitalisierung und Automatisierung im Bahnbereich

Im Rahmen der Digitalisierung und Automatisierung des Schienenverkehrs werden zunehmend große Datenmengen aus unterschiedlichen Quellen aufgezeichnet, beispielsweise strecken- oder fahrzeugseitig erfasste Sensordaten zur Zustandsüberwachung der Schieneninfrastruktur. Diese Daten bilden die Grundlage für bahnbetrieblich und betriebswirtschaftlich wichtige Entscheidungen. Daher müssen immer größer werdende Datenmengen gespeichert, organisiert und überprüft werden, damit die in ihnen enthaltenen Informationen zuverlässig und nachhaltig genutzt werden können.



## 1. Einleitung

Intelligent miteinander verknüpften Daten mit Orts- und Zeitbezug („Smart Big Geodata“) sowie dem aus ihnen mittels Algorithmen abgeleiteten neuen Wissen kommt eine Schlüsselrolle bei der technologischen und wirtschaftlichen Bewältigung der Digitalisierung und Automatisierung des Bahnverkehrs zu. Qualitativ hochwertige und leicht verfügbare Geodaten sind somit entscheidend für die Entwicklung neuer Technologien und die Ertüchtigung der Bahninfrastruktur.

Das Deutsche Zentrum für Luft- und Raumfahrt e.V. (DLR) arbeitet im programmübergreifenden Projekt Digitaler Atlas 2.0 [1] an der Entwicklung einer Geodateninfrastruktur, die alle relevanten Daten für die Entwicklung und das Management des zukünftigen, vernetzten und automatisierten Verkehrs am Boden und zu Wasser bereithält. Ergänzend werden Werkzeugketten entwickelt, um diese Daten zu verarbeiten, zu analysieren, zu bewerten, mit anderen Daten zu verknüpfen und für die weitere Verwendung bereitzustellen.

Das DLR betreibt im Kontext verschiedener nationaler und internationaler Forschungsprojekte eigene fahrzeuggetragene Multi-Sensoreinheiten für die Ortung und Zustandsüberwachung von Schienenfahrzeugen und der Gleisinfrastruktur. Daten dieser Systeme dienen der Entwicklung von Algorithmen für die zustandsbasierte

Instandhaltung [2]. Die im Zuge dieser langjährigen Forschung anfallenden Datenmengen müssen nicht nur übermittelt und gespeichert, sondern auch systematisch und nachvollziehbar aufbereitet und für die weitere Forschungsarbeit bereitgestellt werden.

Im Zuge des Innovationsprogramms 5 Schienengüterverkehr des EU-Förderprogramms für Innovationen im Schienenverkehr, Shift2Rail, hat das DLR-Institut für Verkehrssystemtechnik Werkzeugketten für die automatisierte Erfassung, Konvertierung und Kontrolle von Daten fahrzeugseitiger Sensoren für die Zustandsüberwachung entwickelt [3], die im Weiteren exemplarisch erläutert werden.

Die Gewinnung des in den Daten enthaltenen Wissens stellt sowohl aus wissenschaftlicher und technologischer als auch aus gesellschaftlicher Sicht eine große Herausforderung dar. Das DLR setzt seine Kompetenzen in der Anwendung von Big-Data-Technologien, der Entwicklung von Algorithmen sowie der Softwaremethodik gezielt ein, um zur Lösung dieser Herausforderung beizutragen [4]. Für datenzentrierte Analysetechniken spielt darüber hinaus die gesicherte Qualität der genutzten Daten eine entscheidende Rolle. Im Rahmen der Forschungsaktivitäten des DLR werden, wo immer möglich, quelloffene und weitverbreitete Frameworks verwendet, um zuverlässige, kostengünstige und transferierbare Werkzeugketten zu implementieren.



**Sangeetha Shankar**

Wissenschaftliche Mitarbeiterin Institut für Verkehrssystemtechnik  
Deutsches Zentrum für Luft- und Raumfahrt e.V. (DLR)  
sangeetha.shankar@dlr.de



**Judith Heusel**

Wissenschaftliche Mitarbeiterin Institut für Verkehrssystemtechnik  
Deutsches Zentrum für Luft- und Raumfahrt e.V. (DLR)  
judith.heusel@dlr.de



**Oliver Böttcher**

Wissenschaftlicher Mitarbeiter Institut für Verkehrssystemtechnik  
Deutsches Zentrum für Luft- und Raumfahrt e.V. (DLR)  
oliver.boettcher@dlr.de



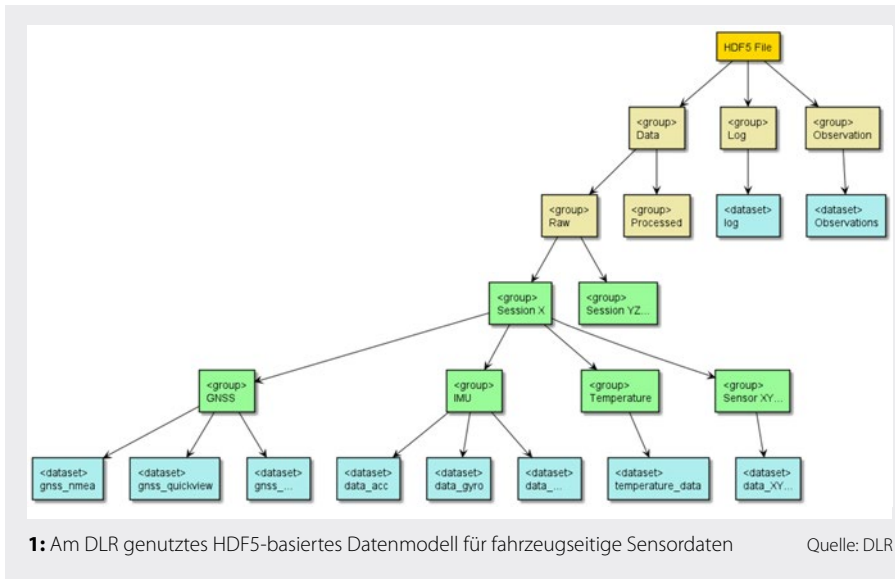
**Akhil Patil**

Wissenschaftlicher Mitarbeiter Institut für Verkehrssystemtechnik  
Deutsches Zentrum für Luft- und Raumfahrt e.V. (DLR)  
akhil.patil@dlr.de



**Benjamin Baasch**

Wissenschaftlicher Mitarbeiter Institut für Verkehrssystemtechnik  
Deutsches Zentrum für Luft- und Raumfahrt e.V. (DLR)  
benjamin.baasch@dlr.de



Der hier vorgestellte Ansatz für das Datenmanagement fahrzeuggetragener Multi-Sensor-Systeme besteht aus drei Schritten:

- Aufzeichnung von Multi-Sensor-Daten unter Verwendung eines gemeinsamen Zeitbezugssystems,
- Konvertierung der Messdaten inklusive aller Metadaten in ein hierarchisches Datenformat mit einem spezifischen Datenmodell und
- Automatisierte Datenkontrolle und Konsistenzprüfung.

**2. Datenformate und -modelle für die Verwaltung von Multi-Sensor-Daten**

Performante Datenformate und -modelle sind eine Voraussetzung für ein effizientes Datenmanagement sowie die Datenanalyse von Multi-Sensor-Daten. Sowohl der Zugriff auf die Daten als auch die Auswertung und Analyse der Daten soll durch eine Standardisierung des Zugriffs vereinfacht werden. Ein Standard-Datenformat sollte Metadaten, die die Parameter und die Konfiguration des Messsystems und der Messumgebung vollständig beschreiben, bereitstellen. Das Datenformat sollte weiterhin strukturiert, flexibel, skalierbar und plattformunabhängig sein. Die Implementierung sollte auf existierenden und gut etablierten Formaten, wie z.B. dem Hierarchical Data Format HDF5, aufbauen.

**2.1. Datenerfassung mit Multi-Sensor-Systemen**

Das DLR-Institut für Verkehrssystemtechnik benutzt für die Erfassung von Daten

mit Multi-Sensoreinheiten das Robot Operating System (ROS) [5]. ROS ist ein quell-offenes Framework, das aus einer Sammlung von Bibliotheken und Werkzeugen zur Implementierung von Applikationen für Roboter und Messsysteme besteht. ROS agiert als ein verteiltes System, bei dem die entwickelten Komponenten lose gekoppelt voneinander arbeiten. Das rosbag, auch als „bag“ bezeichnet, ist das Dateiformat innerhalb von ROS, um die Daten der einzelnen Komponenten aufzunehmen und abzuspeichern. Bei der Datenaufzeichnung werden die Daten durch die einzelnen Komponenten mittels verschiedener sogenannter Topics bereitgestellt. Im Rosbag sind die Daten entsprechend der Topics unterteilt. Die Aufnahme in rosbags erlaubt es, die aufgezeichneten Daten im Nachhinein wieder im ursprünglichen zeitlichen Zusammenhang abzuspielen und somit die Situationen wie zum Aufnahmezeitpunkt wiederzugeben. Die Tools von ROS können dabei helfen, die rosbags nach Topics gefiltert abzuspielen, sodass nur die für den eigenen Verwendungszweck relevanten Daten genutzt werden können.

Einige Vorteile der Nutzung von ROS sind die Quelloffenheit, die große internationale Verbreitung in Wissenschaft und Industrie, die kontinuierliche Weiterentwicklung durch eine lebendige User-Community, die Plattformunabhängigkeit, die Sprachenunabhängigkeit in der Entwicklung, da sowohl C++, aber auch Python genutzt werden können, die Benutzerfreundlichkeit sowie die gute Skalierbarkeit.

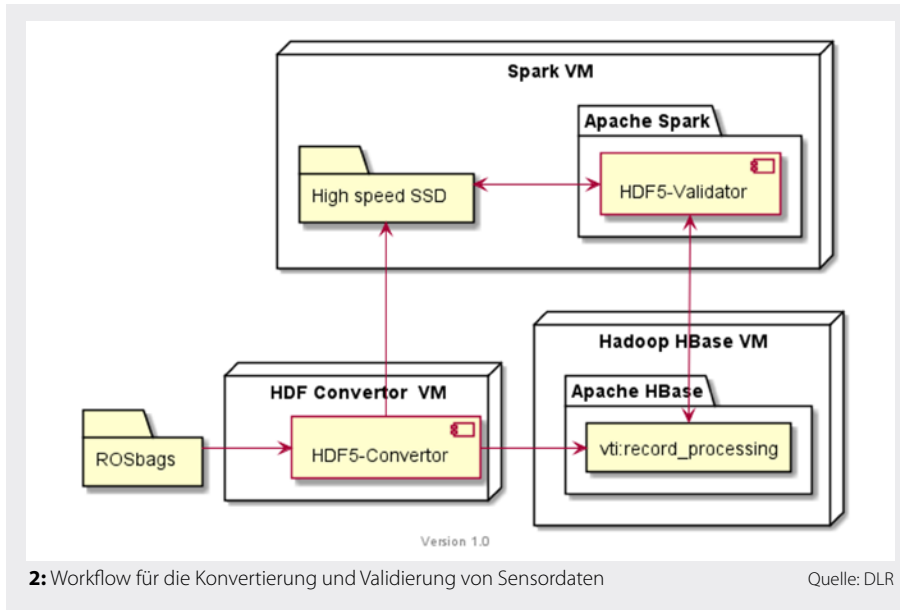
Im hier beschriebenen Anwendungsfall werden Daten aus unterschiedlichen

Quellen, wie z.B. von Empfängern globaler Navigationssatellitensysteme (englisch Global Navigation Satellite System, GNSS), analogen vibro-akustischen Sensoren oder Inertialmesseinheiten mit einem gemeinsamen Zeitbezugssystem in rosbags aufgezeichnet. Dies ermöglicht, Daten mit sehr unterschiedlichen Abtastraten von beispielsweise 1 Hz (GNSS-Empfänger) bis zu zweistelligen kHz (Aufnahmefrequenz der vibro-akustischen Sensoren) zu verknüpfen. Die in Topics strukturierten Daten werden in sogenannte Sessions unterteilt gespeichert. Diese Sessions ergeben sich beim Ein- und Ausschalten des Messsystems und erzeugen jedes Mal ein neues rosbag mit einer eindeutigen Session-ID.

**2.2. HDF5-basiertes Datenmodell für die Bereitstellung und Weiterverarbeitung von Sensordaten**

Das *Hierarchical Data Format* HDF5 [6] ist ein weitverbreitetes und quelloffenes Dateiformat zur Speicherung großer Datenmengen. Ein Vorteil des Formates ist es, dass Daten zusammen mit zugehörigen Metadaten abgelegt werden können. Es unterstützt viele typische Datentypen und ermöglicht das effiziente Lesen und Schreiben von großen Datensätzen. Die HDF5-Programmsammlung besteht aus Bibliotheken in verschiedenen Programmiersprachen zur Handhabung von HDF5-Dateien. Im hier beschriebenen Anwendungsfall werden die in rosbags gesammelten Sensordaten in das HDF5-Format konvertiert, um eine einfache und effiziente Verarbeitung großer Datenmengen mithilfe der HDF5-Bibliotheken zu ermöglichen. Dies erleichtert zum einen das Lesen der Daten in unterschiedlichen Entwicklungssprachen als auch den Austausch und die gemeinsame Nutzung der Daten mit Partnern in einem standardisierten Datenformat. Zudem ist HDF5 auf die Speicherung heterogener Datensätze und insbesondere auf die strukturierte Speicherung der zu den Daten gehörigen Metadaten ausgelegt. Für diesen Zweck wurde das nachstehend vorgestellte Datenmodell definiert.

Das HDF5-Datenmodell besteht aus unterschiedlichen Elementen, um entsprechende Datensätze strukturiert abzubilden. Dies wird hierarchisch abgebildet und besteht aus *groups*, *datasets* und *attributes*. Detaillierte Informationen über das HDF5 Datenmodell sind unter [7] verfügbar. Bild 1 zeigt das HDF5-basierte Datenmodell, wel-



ches speziell für Daten von fahrzeuggetragenen Sensoren am DLR-Institut für Verkehrssystemtechnik entwickelt wurde. Das Modell besteht aus drei Hauptgruppen:

- **„Data“:** Die Hauptgruppe, in der die Datensätze gespeichert werden. Sie hat zwei Untergruppen – „Raw“ und „Processed“ – für die Speicherung von Rohdaten bzw. verarbeiteten Daten.
- **„Log“:** Die Logdateien des Konvertierungsprozesses werden in dieser Gruppe gespeichert.
- **„Observation“:** Diese Gruppe beinhaltet ein Dataset „Observations“, in dem die Metadaten (wie Session-Nummer, Start- und Endzeit der Datenerfassung, verwendete Sensoren) der in der Gruppe „Data“ vorhandenen Datensätze gespeichert werden.

In der Gruppe „Data“ wird je eine Untergruppe pro Session erstellt. Jede dieser Gruppen enthält je eine Untergruppe pro Sensor. Jede sensorspezifische Gruppe enthält einen oder mehrere Datensätze mit den von den Sensoren aufgezeichneten Daten. Das aktuelle Modell unterstützt die Speicherung der Daten von GNSS-Empfängern, inertialen Messeinheiten, Beschleunigungssensoren, Temperatursensoren, Kameras, Luftdruckmessern, Strömungs- und Spannungsmessern, Wettersensoren, Radar und Hodometern. Die Datensätze mit Sensordaten sind Compound-Datasets, d.h., die Datentypen können spaltenabhängig gewählt werden. Alle Datasets haben eine Spalte „timestamp“ für einen Zeitstempel, mit der

mehrere Datensätze anhand der gemeinsamen Zeitbasis zusammengeführt, verglichen und analysiert werden können.

### 3. Werkzeugkette zur Datenkonvertierung

Bild 2 zeigt einen Überblick über den Workflow der Konvertierung und Validierung von Messdaten. Der Konvertierungsprozess ist in diesem Abschnitt erläutert; die Validierung wird im nächsten Abschnitt dargestellt. Für diese Prozessschritte wird aufgrund der großen Datenmengen ein hoch-performerter und mit entsprechenden quelloffenen Frameworks (Apache

Spark, Apache Hadoop etc.) aufgebauter Rechnerverbund des DLR verwendet. Die Architektur dieses Rechnerverbundes wird in [8, 9] erläutert.

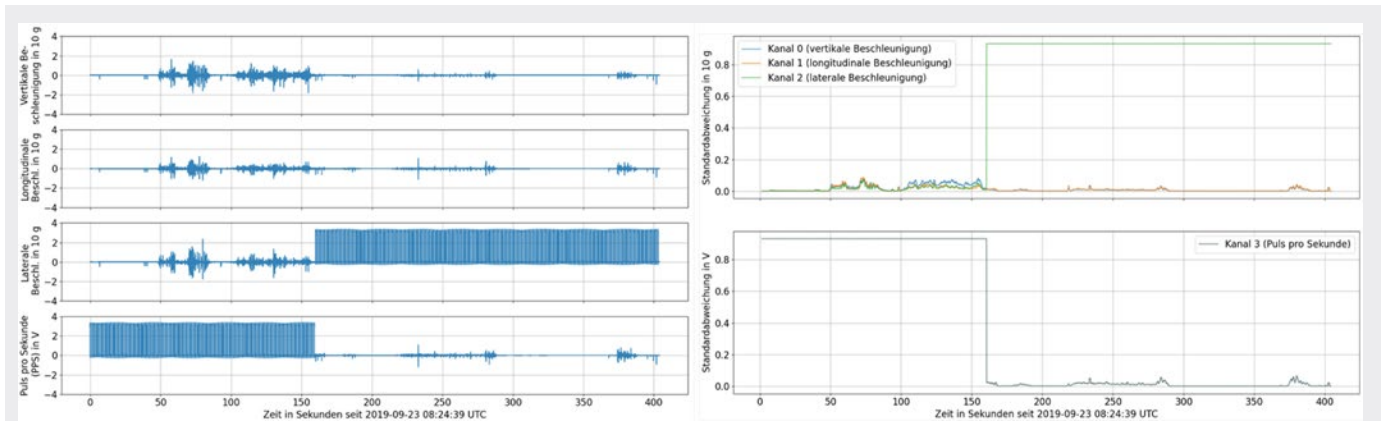
Die Rohdaten der Sensoren werden als rosbags von den fahrzeuggetragenen Messsystemen in eine Speicherinfrastruktur des Rechnerverbundes umgelagert. Für die automatische Umwandlung der rosbags in das HDF5-Format wurde ein entsprechender Konverter implementiert. Die aus der Konvertierung resultierenden HDF5-Dateien werden im nächsten Schritt auf einem Hochgeschwindigkeitsspeicher des Rechnerverbundes abgelegt, um einen performanten Zugriff zu gewährleisten. Im beschriebenen Anwendungsfall variiert die Größe der Dateien von einigen hundertn Megabytes bis zu mehreren Gigabytes. Der Speicherbedarf ist dabei abhängig von der Laufzeit der Sessions und der Anzahl der aktiven Sensoren. Ein Beispiel einer konvertierten HDF5-Datei ist in Bild 3 dargestellt.

Der Status der Konvertierungen wird regelmäßig in einer HBase-Tabelle protokolliert. Apache HBase [10] ist eine verteilte Datenbank speziell für die Speicherung von sehr großen Datenmengen bzw. zahlreichen Tabelleneinträgen. Insbesondere bei einer sehr großen und ständig wachsenden Anzahl von HDF5-Dateien im Laufe der Zeit ist ein entsprechend effizientes System für die Verwaltung unerlässlich. Der Konverter erzeugt pro Session einen Eintrag (Zeile) in der Tabelle. Die Tabelle dient der Speicherung aller relevanten Details der einzelnen zeitgestempelten Konvertie-

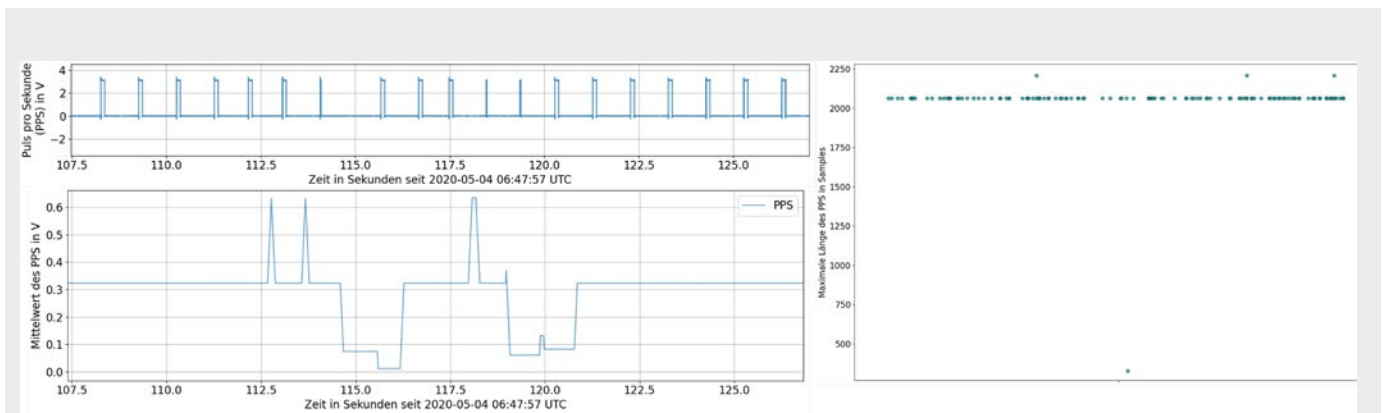
	timestamp	temperature	variance
0	1455148804436947107	21.9375	0.0
1	1455148800001471042	21.9375	0.0
2	1455148800002906084	21.9375	0.0
3	1455148800004400014	21.9375	0.0
4	14551488000080710887	21.9375	0.0
5	14551488000083597898	21.9375	0.0
6	14551488000087310075	21.9375	0.0
7	14551488000089155912	21.9375	0.0
8	14551488000090826988	21.9375	0.0
9	14551488000092508077	21.9375	0.0
10	14551488000094217061	21.9375	0.0
11	14551488000095854043	21.9375	0.0
12	14551488000097415924	21.9375	0.0
13	1455148800162851095	21.9375	0.0
14	1455148800165532112	21.9375	0.0
15	1455148800167787075	21.9375	0.0
16	1455148800170118093	21.9375	0.0

3: Beispiel einer konvertierten HDF5-Datei

Quelle: DLR



4: Beispiel für eine fehlerhafte Datenaufnahme: vertauschte Aufnahmequellen ab ca. Sekunde 160 auf Kanal 2 und 3 des AD-Wandlers (links, Reihe 3 und 4). Die oberen beiden Reihen zeigen die Kanäle 1 und 2. Rechts: Laufende Standardabweichung der Daten der verschiedenen Kanäle mit einer Fensterlänge von einer Sekunde  
Quelle: DLR



5: Beispiel für eine fehlerhafte Datenaufnahme (fehlende Datenpakete). Links oben: Puls pro Sekunde. Sichtbar sind zu kurze Pulse bei Sekunde 114, 118 und 119; fehlender Puls bei Sekunde 115; zu kurze Abstände bei Sekunde 112, 113 und 117. Links unten: Laufender Mittelwert des PPS über die vorangegangene Sekunde. Der Wert sollte konstant 0,33 V betragen und weicht an den betroffenen Stellen stark ab. Rechts: Maximale Länge des PPS in Samples pro Session für den betroffenen AD-Wandler für 123 Sessions. Der Idealwert beträgt 2062,5 Messwerte. Die betroffene Session weicht mit einem Wert von 2208 stark ab  
Quelle: DLR

rungs- und Validierungsschritte einschließlich aufgetretener Warnungen oder Fehler. Die erfassten Informationen ermöglichen so eine vollständige Nachvollziehung aller Verarbeitungsschritte und die Überwachung des gesamten Prozesses. Der Status des gesamten Rechnernetzes und aller laufender Prozesse wird ergänzend durch Prometheus [11] und Grafana [12] kontinuierlich überwacht.

4. Qualitätskontrolle von Sensordaten

Um die Daten des Multi-Sensor-Systems gewinnbringend und zuverlässig einsetzen zu können, dürfen nur Daten in einwandfreier Qualität weiterverarbeitet und für die Datenanalyse – beispielsweise zur Gewinnung von Informationen über den Zustand eines Assets oder zur Bestimmung der Position eines Schienenfahrzeuges – verwendet werden. Typische Ursa-

chen für eine geringe Datenqualität sind beispielsweise defekte oder gelöste Sensoren, beschädigte Kabel oder eine instabile Spannungsversorgung. Daher müssen die erfassten Sensordaten an verschiedenen Punkten der Prozessierungskette automatisiert überprüft werden. Die Überprüfung umfasst einerseits eine Plausibilitätsprüfung der Sensorrohdaten, insbesondere auf untypische Werte, Widersprüche zwischen den Daten unterschiedlicher Sensoren sowie auf fehlende Daten; andererseits den Vergleich verschiedener Parameter vor und nach den Verarbeitungsschritten, um Fehler während der Umwandlung zu erkennen. Neben der Qualitätskontrolle der Daten sowie dem Verhindern von Prozessierungsfehlern dient die Überprüfung auch der Erkennung von Unregelmäßigkeiten bei der Aufnahme der Daten, sodass Störungen bzw. Defekte an den Messsystemen möglichst frühzeitig aufgedeckt

werden können. Zur Verarbeitung der vorhandenen großen Datenmengen (mehrere GB/Tag/Messsystem) bietet sich die Parallelisierung der Berechnung von Parametern mithilfe eines Rechnernetzes (z. B. mit Apache Spark [13]) an.

Die Qualitätskontrolle umfasst zunächst Plausibilitätsprüfungen hinsichtlich einfach zu erfassender Dateneigenschaften. Beispielsweise werden Parameter bezüglich der Zeitstempel der verschiedenen Sensoren (Start, Ende, Zeitspanne, Anzahl der Datenpunkte, Differenzen der Zeitstempel verschiedener Datenpakete) auf Inkonsistenzen überprüft. Neben der Kontrolle auf Abweichungen zwischen Zeitstempeldifferenzen und der erwarteten tatsächlichen Anzahl von Datenpaketen innerhalb eines Aufnahmezeitraums werden die Aufnahmeintervalle der verschiedenen Sensoren abgeglichen. Fehlende Daten oder Datenpakete können so leicht

Dieses Projekt wurde vom Shift2Rail Joint Undertaking (JU) unter der Fördervereinbarung (Grant Agreement) Nr. 826206, Nr. 881778 und Nr. 101004051 im Rahmen des EU-Förderprogramms für Forschung und Innovation Horizont 2020 gefördert.



auffindig gemacht und Ausfälle am Sensorsystem schnell behoben werden.

Darüber hinaus werden pro Datenaufzeichnung verschiedene Parameter wie Mittelwert, Maximum, Minimum und Standardabweichung aus den unterschiedlichen Sensorausgaben extrahiert. Diese müssen vor und nach der Formatkonvertierung übereinstimmen. Die berechneten Parameter werden darüber hinaus zur Qualitätskontrolle verwendet. So können fehlende Datenpakete beispielsweise anhand von Unregelmäßigkeiten eines hochfrequent abgetasteten Sekundensignals (Pulse per Second, PPS) – des GNSS-Empfängers erkannt werden. Im Gegensatz zur Ausgabe des Sekundensignals, dessen erwartete Werte bekannt und Ausreißer daher leicht erkennbar sind, variieren z. B. die Achslagerbeschleunigungen stark in Abhängigkeit vom Bewegungszustand des Schienenfahrzeuges (Stillstand vs. Fahrt, Geschwindigkeit) und der Beschaffenheit der befahrenen Strecken. In diesem Fall können die Sensordaten zur gegenseitigen Überprüfung genutzt werden, beispielsweise zwischen den verschiedenen Messrichtungen (bei triaxialen Beschleunigungssensoren) oder von verschiedenen Beschleunigungssensoren (z. B. den Sensoren von linker und rechter Fahrzeugseite).

Die Bilder 4 und 5 zeigen zwei Beispiele für Unregelmäßigkeiten an einem Messsystem, die über die Qualitätskontrolle entdeckt wurden, sowie im zweiten Fall ein Beispiel für einen Parameter, über den der Fehler detektiert werden kann. Im ersten Fall (Bild 4) wurden innerhalb einer Session die Daten der Kanäle 2 und 3 vertauscht

aufgezeichnet, d. h. das PPS-Signal wurde dort übermittelt, wo die laterale Beschleunigung erwartet wurde und umgekehrt. Diese Tatsache ist sowohl in einer abrupten Änderung des laufenden Mittelwerts bzw. Standardabweichung der Daten der beiden Kanäle, dem Verhältnis der Standardabweichung der zu den Achslagerbeschleunigungsdaten zugehörigen Kanälen sowie im Mittelwert und der Standardabweichung des Kanals 3 (PPS) der gesamten Session im Vergleich zum Normalverhalten erkennbar. Bild 4 zeigt einen Ausschnitt aus einer Session, in dem nicht alle Datenpakete aufgezeichnet wurden. Dies ist ebenfalls in verschiedenen Kennzahlen zu erkennen, beispielsweise in der maximalen und mittleren Länge der PPS-Signale der Session oder lokalisierbar im laufenden Mittelwert über das PPS-Signal, der über 20625 Datenpunkte nahezu konstant nahe 0.33V (Eingangssignal 3.3V  $\pm$  10 %) liegen sollte.

## 5. Fazit und Ausblick

Fahrzeug- und streckenseitige Sensoren werden bereits heute großflächig im Bahnbereich eingesetzt. In der Zukunft wird ihre Bedeutung für einen digitalisierten und automatisierten Schienenverkehr weiter steigen. Dabei fallen immer größer werdende Mengen an Sensordaten mit Raum- und Zeitbezug an. Dem Management und der Qualitätskontrolle dieser großen Mengen von Geodaten kommt dabei eine besondere Rolle zu, um aus den Daten zuverlässige Informationen zu gewinnen und nachhaltig verfügbar zu machen. Das DLR entwickelt zu diesem Zweck entsprechende Datenmodelle und -formate sowie Werkzeugketten. Diese basieren auf gängigen flexiblen, skalierbaren und plattformunabhängigen Big-Data-Technologien bzw. Frameworks für eine verteilte Prozessierung von Daten. Durch die Nutzung von quelloffenen Technologien sowie die Nutzung von standardisierten Schnittstellen, Datenformaten und Datenmodellen können Forschungsaktivitäten und Ergebnisse ziel führend mit Systemen bei Praxispartnern verknüpft werden. Auf diese Weise können neue Ansätze und Konzepten für föderierte Daten- und Dienstplattformen in operativer Umgebung praktisch erprobt werden. Entsprechende Forschungsaktivitäten des DLR sind ein Schwerpunkt von laufenden und zukünftigen kooperativen Forschungsprojekten im Bereich Digitalisierung des Verkehrssystems Bahn. ●

## Literatur

- [1] Digitaler Atlas 2.0. URL <https://verkehrsforchung.dlr.de/de/projekte/digitaler-atlas-20-domaeneneuebergreifende-softwareanwendungen-und-geodateninfrastrukturen> [letzter Zugriff am 29.09.2022].
- [2] Baasch, B.; Heusel, J.; Groos, J. C.; Shankar, S.: Eingebettete Zustandsüberwachung der Gleisinfrastruktur: Entwicklung und Erprobung von eingebetteten Multi-Sensor-Systemen für die kontinuierliche Zustandsüberwachung der Gleisinfrastruktur im operativen Betrieb. EI – DER EISENBAHNINGENIEUR, 12/2019, S. 6–8. URL <https://elib.dlr.de/129363/>.
- [3] Bänsch, R.; Weidner, M.; Mendizabal, J.; Galipienzo, J.; Henning, A.; Baasch, B.; Schmauder, N.; Winkler-Höhn, R.; Gomes Alves, C.; Krüger, D.; Laporte, M.: CFW – Schienengüterverkehrskonzept von morgen. EI – DER EISENBAHNINGENIEUR, 9/2022, S. 138-145.
- [4] Big und Smart Data – DLR Portal. URL <https://www.dlr.de/content/de/artikel/digitalisierung/big-und-smart-data.html> [letzter Zugriff am 04.10.2022].
- [5] ROS/Introduction - ROS Wiki. URL <http://wiki.ros.org/ROS/Introduction> [letzter Zugriff am 29.08.2022].
- [6] HDF5. URL <https://portal.hdfgroup.org/display/HDF5/HDF5> [letzter Zugriff am 18.08.2022].
- [7] HDF5 User Guide. URL [https://portal.hdfgroup.org/display/HDF5/HDF5+User+Guides?previe=53610087/53610088/Users\\_Guide.pdf](https://portal.hdfgroup.org/display/HDF5/HDF5+User+Guides?previe=53610087/53610088/Users_Guide.pdf) [letzter Zugriff am 19.08.2022].
- [8] Shankar, S.; Schubert, L. A.; Patil, A. J.; Erdmann, J.: Der Einsatz von Big Data und der Verkehrssimulation SUMO im Rahmen von Rail2X. SIGNAL+DRAHT, 10/2020, S. 49–58, URL <https://elib.dlr.de/134134/>.
- [9] Shankar, S.; Grünhäuser, M.: Managing digital level crossing data using open-source big data technologies. SIGNAL+DRAHT, 2/2022, S. 31–37. URL <https://elib.dlr.de/140078/>.
- [10] Apache HBase™. URL <https://hbase.apache.org/> [letzter Zugriff am 04.10.2022].
- [11] Prometheus. URL <https://prometheus.io/> [letzter Zugriff am 29.09.2022].
- [12] Grafana. URL <https://grafana.com/> [letzter Zugriff am 29.09.2022].
- [13] Apache Spark™ - Unified Engine for large-scale data analytics. URL <https://spark.apache.org/> [letzter Zugriff am 29.08.2022].

## Summary

### Management of big sensor data amounts for digitization and automation in the rail sector

Sensor data which are being collected in ever-increasing quantities in the course of digitization and automation of rail transport require an appropriate data management. In this article, data formats, -models and tool chains are exemplary presented enabling to record, save, prepare and verify data via multi-sensor units. They build the basis for the development and implementation of algorithms for a data centered information acquisition in the rail sector.