

DaFuER:

# Forschungsprojekt zur Steigerung der Datenqualität

... und welche Rolle die Datenfusion dabei spielt

Mit dem Forschungsprojekt ‚DaFuER‘ sollten, mit Blick auf betriebliche Rückmelde-  
daten, Methoden der Datenfusion zur Steigerung der Datenqualität unter die Lupe  
genommen werden. Lesen Sie hier die wichtigsten Ergebnisse, die am Anwen-  
dungsfall in der Demonstrationsfabrik Aachen veranschaulicht werden. >



DaFuER:

# Research Project for the Increase the Data Quality

... and what role data fusion plays in this process

With the research project 'DaFuER', methods of data fusion for increasing data quality were to be examined with a view to operational feedback data. Here you can read the most important results, which are illustrated by the use case in the *demonstration factory in Aachen*. >

**A**ktuelle Megatrends wie Globalisierung und Digitalisierung fordern produzierenden Unternehmen einiges ab: Eine echtzeitfähige und effiziente Produktionsplanung und -steuerung (PPS) als Basis einer fundierten und kurzfristigen Entscheidungsfindung gehört dazu. Notwendige Bedingungen für eine zuverlässige Entscheidungsfindung sind nach wie vor eine hohe Datenqualität und die anwendungsgerechte Informationsverfügbarkeit. Die zentrale Herausforderung bei der Steigerung der Datenqualität besteht in den Investitionskosten für die Implementierung von entsprechenden Maßnahmen. Besonders betroffen von einer eingeschränkten Investitionsfähigkeit sind kleine und mittlere Unternehmen (KMU).

Eine Möglichkeit zur Senkung der Investitionskosten besteht darin, die Datenqualität eines Datensatzes zu erhöhen, indem man die Daten fusioniert oder verschiedene Datenquellen kombiniert. Das Forschungsprojekt ‚DaFuER‘ knüpfte genau an diesem Punkt an. Ziel war es, die Datenqualität insbesondere für die Produktionssteuerung und das Produktionscontrolling mithilfe der Anwendung von Methoden der Datenfusion sicherzustellen. Dabei wurde im Projekt ein Leitfaden zur Anwendung der Datenfusion im Kontext betrieblicher Rückmeldedaten erarbeitet, der als Entscheidungshilfe für Anwenderunternehmen und für Entwickler:innen betrieblicher Anwendungssysteme dient.

Der Aufbau des Leitfadens ist an dem im Forschungsprojekt entwickelten Modell zur Anwendung der Datenfusion im Kontext betrieblicher Rückmeldedaten ausgerichtet.

In einem ersten Arbeitsschritt definieren die Anwendenden den aktuellen Anwendungsfall. Dabei ermitteln sie zunächst die wichtigsten Informationsbedarfe; dies geschieht auf Basis einer Übersicht, die die für die PPS relevanten Informationen zeigt. Die Anwendenden wählen diejenigen Informationen aus, die sich für sie aufgrund einer mangelhaften Datenqualität als unzuverlässig erweisen. Die Informationsverfügbarkeit ermitteln sie auf Basis der verfügbaren Daten. Diese ergeben sich aus den dem Unternehmen zur Verfügung stehenden Datenquellen.

Im zweiten Arbeitsschritt werden die zu fusionierenden Datenquellen bestimmt. Für jede als relevant identifizierte Information werden in einer Übersicht potenziell dazugehörige Datenquellen aufgelistet, aus denen sich diese Informationen extrahieren lassen. Die Anwendenden können dadurch ihrem Informationsbedarf konkrete, verfügbare Datenquellen zuordnen. Darüber hinaus ermitteln sie qualitativ für jede Datenquelle, in welchem Maß diese die verschiedenen Kriterien der Datenqualität erfüllt. Auf Basis dieser Qualitätsmerkmale ist es den Anwendenden möglich, diejenigen verfügbaren Datenquellen für die Fusion auszuwählen, die, sich gegenseitig ergänzend, die Datenqualitätsmerkmale erfüllen.

Schließlich werden die geeigneten Methoden der Datenfusion ermittelt. Auf Basis einer Klassifikation der betrachteten

**C**urrent megatrends such as globalization and digitization make great demands on manufacturing companies: Real-time and efficient production planning and control (PPC) as the basis for sound and short-term decision-making is one of them. Necessary conditions for reliable decision-making continue to be high data quality and application-oriented information availability. The central challenge in increasing data quality lies in the investment costs for implementing appropriate measures. Small and medium-sized enterprises (SMEs) are particularly affected by a limited ability to invest.

One way to reduce investment costs is to increase the data quality of a dataset by merging data or combining different data sources. The research project ‘DaFuER’ took up exactly this point. The aim was to ensure data quality, especially for production management and production controlling, by applying data fusion methods. The project developed a guideline for the application of data fusion in the context of operational feedback data, which serves as a decision-making aid for user companies and developers of operational application systems.

The structure of the guide is based on the model developed in the research project for the application of data fusion in the context of operational feedback data. In a first step, the users define the current use case. They first determine the most important information needs; this is done on the basis of an overview showing the information relevant for PPS. The users select the information that they consider unreliable due to poor data quality. They determine the information availability based on the available data. These are derived from the data sources available to the company.

In the second step, the data sources to be merged are determined. For each piece of information identified as relevant, an overview lists potentially related data sources from which this information can be extracted. Users can thus assign concrete, available data sources to their information needs. In addition, they qualitatively determine for each data source to what extent it fulfills the various data quality criteria. Based on these quality characteristics, users are able to select those available data sources for fusion that, complementing each other, fulfill the data quality characteristics.

Finally, the appropriate methods of data fusion are identified. Based on a classification of the considered data sources, the users first developed a morphology to describe a data source generically. By combining all possible types of data sources or their morphological characteristics, process-typical errors during fusion can be derived and error classes can be formed. During the project, it was determined to what extent a method for applying data fusion is resistant to a corresponding type of error. If it is known which data sources are to be fused, it is possible for the users to derive concrete methods of data

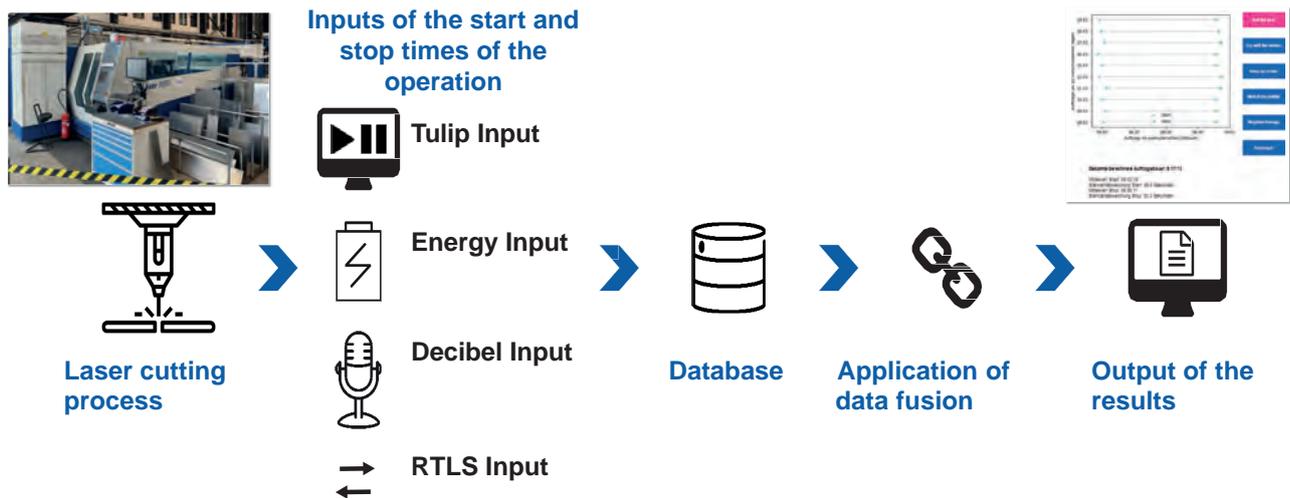


Figure 1 : Use Case DFA

Datenquellen entwickelten die Anwendenden zunächst eine Morphologie, um eine Datenquelle generisch beschreiben zu können. Indem alle möglichen Arten von Datenquellen bzw. deren morphologische Ausprägungen kombiniert werden, können prozesstypische Fehler bei der Fusion abgeleitet und Fehlerklassen gebildet werden. Während des Projekts wurde bestimmt, inwieweit eine Methode für die Anwendung der Datenfusion resistent gegen einen entsprechenden Fehlertyp ist. Wenn bekannt ist, welche Datenquellen fusioniert werden sollen, ist es den Anwendenden auf Basis dieser Zuordnung möglich, konkrete Methoden der Datenfusion abzuleiten. Jene können somit die zentralen Herausforderungen bei der Kombination der gewählten Datenquellen bestmöglich bewältigen.

Um das in der Theorie entwickelte Modell in der praktischen Anwendung zu prüfen, wurde unter anderem ein Usecase in der Demonstrationsfabrik Aachen (DFA) durchgeführt.

Hintergrund dieses Usecase waren Unregelmäßigkeiten beim Vergleich von Start- und Endzeitpunkten einzelner Arbeitsgänge einer Laserschneidmaschine. Die vorliegende Datenqualität ist aktuell insbesondere unter den Aspekten *Genauigkeit*, *Vollständigkeit* und *Aktualität* mangelhaft.

Ziel der praxisnahen Umsetzung war es, die Datenqualität zu verbessern, indem die verschiedenen Informationsquellen mithilfe von Datenfusionsmethoden kombiniert werden.

Der zu optimierende Informationsbedarf betraf die Arbeitsgangzeit der Laserschneidmaschine. Die Informationsverfügbarkeit bestand in den erfassten Daten eines Buzzers, eines RTLS-Sensors, einer Intelligenten Sensorik und eines ergänzenden Informationssystems. Die Intelligente Sensorik enthielt einen Energie- und einen Noise-Sensor. Die Sensoren sendeten das Signal „Start“ bzw. „Stopp“, wenn der Energieverbrauch bzw. der Geräuschpegel eine bestimmte Schwelle über- bzw. wieder unterschritt. Der Buzzer sendete entsprechende Signale bei ma-

fusion based on this assignment. Those can thus best address the key challenges in combining the selected data sources. In order to test the model developed in theory in practical application, a use case was carried out at the Demonstration Factory Aachen (DFA), among other things.

The background to this use case was irregularities in the comparison of start and end times of individual operations of a laser cutting machine. The existing data quality is currently inadequate, especially with regard to accuracy, completeness and up-to-dateness.

The goal of the practical implementation was to improve the data quality by combining the different information sources using data fusion methods.

The information need to be optimized was related to the laser cutting machine operation time. The information availability consisted of the acquired data from a buzzer, an RTLS sensor, an Intelligent Sensor Technology, and a complementary information system. The Intelligent Sensor System included an energy sensor and a noise sensor. The sensors sent the “start” and “stop” signals, respectively, when the energy consumption or noise level exceeded or fell below a certain threshold. The buzzer sent corresponding signals when the employee performed manual actions. The RTLS sensor was able to measure when employees entered or left a predefined work zone.

In order to select the best possible data fusion method for the application, it was first necessary to identify potentially occurring defect classes. For this purpose, the morphological characteristics of the data sources were combined. Thus, with regard to the use case, potentially large data volumes, the lack of possibility to apply mathematical operators, and different scale levels, among others, could be identified as errors typical for the process. Based on the analyzed resistances of the data fusion methods to the collected error classes, the method

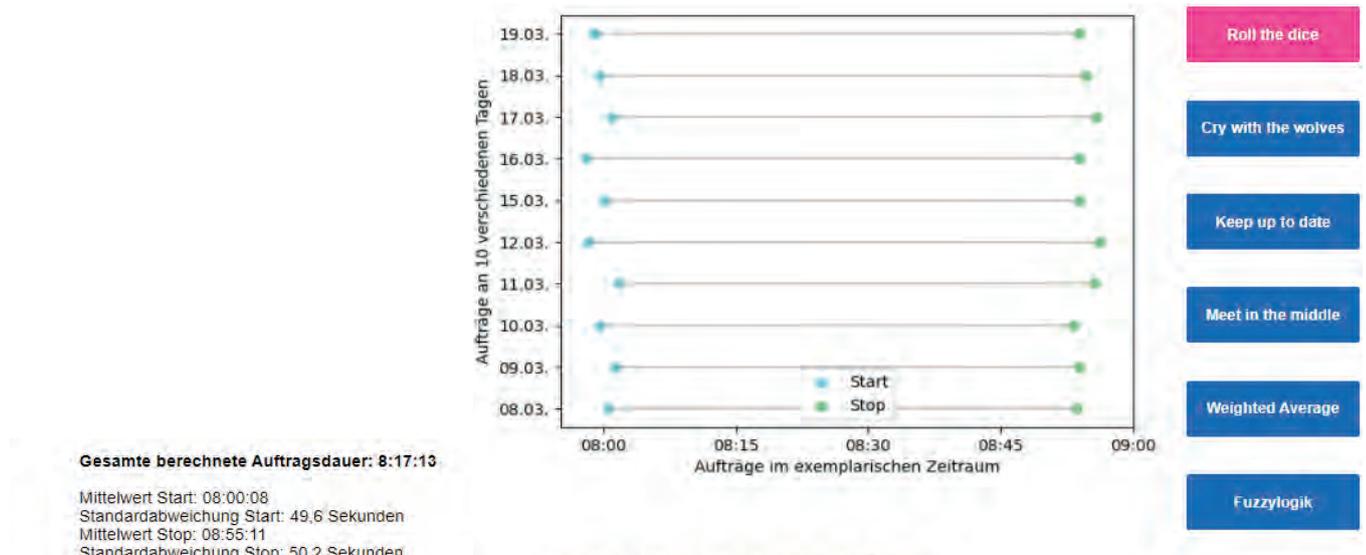


Figure 2: Result of the data fusion (Roll the Dice)

nueller Tätigkeit des Mitarbeitenden. Über den RTLS-Sensor konnte gemessen werden, wann die Mitarbeitenden eine fest definierte Arbeitszone betraten bzw. wieder verließen.

Um für den Anwendungsfall die bestmögliche Datenfusionsmethode auszuwählen, mussten zunächst potenziell auftretende Fehlerklassen identifiziert werden. Dafür wurden die morphologischen Ausprägungen der Datenquellen miteinander kombiniert. So konnten im Hinblick auf den Usecase u. a. potenziell große Datenmengen, die fehlende Möglichkeit zur Anwendung mathematischer Operatoren und unterschiedliche Skalenniveaus als prozesstypische Fehler ausgemacht werden. Basierend auf den analysierten Resistenzen der Datenfusionsmethoden gegenüber den gesammelten Fehlerklassen, eigneten sich im Rahmen dieses Usecase insbesondere die Methodenkategorien *Entscheidungsregeln*, *Fuzzylogik* und *Künstliche Neuronale Netze* zur Fusion der erfassten Daten.

Die entsprechenden Methoden wurden daraufhin hinsichtlich ihrer Anwendbarkeit geprüft. Da es sich, basierend auf dem Fertigungskonzept bei den Arbeitsgängen an der Laserschneidanlage, um sehr viele unterschiedliche Aufträge handelte, wurden Methoden, die eine große Datengrundlage erfordern (z. B. Künstliche Neuronale Netze), ausgeschlossen. Auf dieser Grundlage wurden im genannten Anwendungsfall verschiedenste Entscheidungsregeln sowie eine Methodik der Fuzzylogik implementiert.

Figure 2 zeigt eine Auswertung der generierten Daten anhand der Datenfusionsmethode, genauer gesagt der Entscheidungsregel, „Roll the Dice“. Dargestellt sind die mit der genannten Methode generierten Start- und Stopp-Zeitpunkte für zehn verschiedene Tage im selben Zeitraum. Darüber hinaus werden die gesamte Auftragsdauer über die

categories decision rules, fuzzy logic, and artificial neural networks were particularly suitable for the fusion of the collected data in the context of this usecase.

The corresponding methods were then tested with respect to their applicability. Since, based on the manufacturing concept, the operations on the laser cutting system involved a very large number of different jobs, methods requiring a large data base (e.g. artificial neural networks) were excluded. On this basis, a wide variety of decision rules and a fuzzy logic methodology were implemented in the aforementioned use case.

Figure 2 shows an evaluation of the generated data using the data fusion method, more precisely the decision rule, “Roll the Dice”. Shown are the start and stop times generated using the aforementioned method for ten different days in the same time period. In addition, the total order duration over the ten days as well as mean values and standard deviations of the start and stop times are shown for comparison purposes. Thus, for the aforementioned decision rule, the calculated total job duration is approximately 8 hours and 17 minutes. Depending on the data fusion method used, different order durations or mean values and standard deviations result. The results of the application of further data fusion methods for this use case can be found on the website of the research project<sup>1</sup>.

The research project provided users with two options for the actual implementation of data fusion: one option is to step through the process steps of the data fusion application model using the application guide. An alternative is the free, interactive online tool created as part of this research project,

<sup>1</sup> dafuer-tool.fir.de

zehn Tage sowie Mittelwerte und Standardabweichungen der Start- und Stopp-Zeitpunkte zu Vergleichszwecken dargestellt. So ergibt sich für die genannte Entscheidungsregel eine berechnete Gesamt-Auftragsdauer von ca. 8 Stunden und 17 Minuten. In Abhängigkeit der verwendeten Datenfusionsmethode ergeben sich unterschiedliche Auftragsdauern bzw. Mittelwerte und Standardabweichungen. Die Ergebnisse der Anwendung weiterer Datenfusionsmethoden für diesen Anwendungsfall können auf der Website des Forschungsprojekts<sup>1</sup> nachvollzogen werden.

Das Forschungsprojekt bot den Anwendenden zwei Möglichkeiten zur konkreten Umsetzung der Datenfusion: Eine Möglichkeit ist das schrittweise Durchlaufen der einzelnen Prozessschritte des Modells zur Anwendung der Datenfusion mithilfe des Anwendungsleitfadens. Eine Alternative dazu bietet das im Rahmen dieses Forschungsprojekts erstellte kostenlose, interaktive Online-Tool, in dem die Anwendenden Schritt für Schritt durch die Thematik geführt und angeleitet werden.

Die Projektergebnisse ermöglichen ganz konkret kleinen und mittleren Unternehmen, in der operativen Auftragsdurchführung die Informationstransparenz zu erhöhen. Dies gelang, indem Qualitätsanforderungen fundiert und präzise definiert wurden. Zudem konnte man die Datenqualität erheblich steigern, indem man die entwickelten Lösungen anhand ausgewählter Beispiele in der Praxis anwendete. Hierdurch wurden die Entscheidungsprozesse in der Produktionssteuerung und im Produktionscontrolling, aber auch zur Kennzahlenerhebung vereinfacht.

which guides and instructs users step-by-step through the topic. The project results make it possible for small and medium-sized enterprises to increase information transparency in the operational execution of orders. This was achieved by defining quality requirements in a well-founded and precise manner. In addition, it was possible to significantly increase data quality by applying the developed solutions in practice using selected examples. This simplified the decision-making processes in production management and production controlling, as well as for the collection of key figures.

jn

#### Literatur

DIENES, C.; PAHNKE, A.; WOLTER, H.-J.: *Investitionsverhalten von kleinen und mittleren Unternehmen*. IfM-Materialien; Nr. 268. Institut für Mittelstandsforschung (IfM) Bonn, März 2018. [https://www.ifm-bonn.org/fileadmin/data/redaktion/publikationen/ifm\\_materialien/dokumente/IfM-Materialien-268\\_2018.pdf](https://www.ifm-bonn.org/fileadmin/data/redaktion/publikationen/ifm_materialien/dokumente/IfM-Materialien-268_2018.pdf) (Link zuletzt geprüft: 27.07.2021)

RUSER, H.; PUENTE LEÓN, F.: *Informationsfusion – Eine Übersicht (Information Fusion – An Overview)*. In: *tm – Technisches Messen* 74 (2007) 3, S. 93 – 102. DOI: 10.1524/teme.2007.74.3.93.

SCHUH, G.; REUTER, C.; BRAMBRING, F.; PROTE, J.-P.; HEMPEL, T.; GÜTZLAFF, A.: *Organisation und IT*. In: *Handbuch Industrie 4.0. Geschäftsmodelle, Prozesse, Technik*. Hrsg.: G. Reinhart. Hanser, München [u. a.] 2017, S. 146–154.

jn

<sup>1</sup> [dafuer-tool.fir.de](http://dafuer-tool.fir.de)

**Project Title:** DaFuER

**Funding/Promoters:** Bundesministerium für Wirtschaft und Energie (BMWi);  
Arbeitsgemeinschaft industrieller Forschungsvereinigungen „Otto von Guericke“ e. V. (AiF)

**Funding no.:** 20579 N

**Project Partner:** AUTO HEINEN GmbH; Berghof Systeme e.K.; DFA Demonstrationsfabrik Aachen GmbH;  
INDUTRAX GmbH; Maschinenfabrik Möllers GmbH; Mattern Consult Gesellschaft für  
Produktionsregelung und Logistik mbH; mk Plast GmbH & Co. KG; NETRONIC Software GmbH;  
SICK AG; Ubisense AG; Westaflexwerk GmbH

**Website:** [dafuer.fir.de](http://dafuer.fir.de)

The IGF project 20579N of the research association FIR e. V. at RWTH Aachen University was funded by the Federal Ministry for Economic Affairs and Energy (BMWi) via the AiF within the framework of the program for the promotion of joint industrial research (IGF) on the basis of a resolution of the German Bundestag.



Jokim Janßen, M.Sc.  
Project Manager  
Production Management  
FIR e. V. at RWTH Aachen University  
Phone: +49 241 47705-413  
Email: [Jokim.Janssen@fir.rwth.aachen.de](mailto:Jokim.Janssen@fir.rwth.aachen.de)

