www.fir.rwth-aachen.de

LEITFADEN

# Anwendung der Datenfusion bei der Erfassung und Speicherung betrieblicher Rückmeldedaten (DaFuER)

Jokim Janßen · Tobias Schröer · Tobias Wagner

fir an der RWTH Aachen

# Impressum

Autoren:
Jokim Janßen · FIR e. V. an der RWTH Aachen
Tobias Schröer · FIR e. V. an der RWTH Aachen
Tobias Wagner· FIR e. V. an der RWTH Aachen

Bildnachweise:
Titelbild: © Epstudio20 – stock.adobe.com; S.4: ©peterschreiber.media – stock.adobe.com;
S. 10: © peshkova – stock.adobe.com; S. 12: © Sergey Nivens – stock.adobe.com

# Lizenzbestimmungen/Copyright


FIR e. V. an der RWTH Aachen
Campus-Boulevard 55
52074 Aachen
Tel.: +49 241 47705-0
E-Mail: info@fir.rwth-aachen.de
www.fir.rwth-aachen.de

# Inhaltsverzeichnis

# Management-Summary

Produzierende Unternehmen sind heutzutage aufgrund zunehmender Konkurrenz aus Niedriglohnländern und eines schrumpfenden Technologievorsprungs einem enormen Kostendruck ausgesetzt, sodass Konzepte zur Steigerung der Produktivität erforderlich werden. Diese Konzepte sind vor allem auf die Optimierung innerbetrieblicher Abläufe auf Basis von Rückmeldedaten ausgerichtet. Eine notwendige Bedingung für das Ausschöpfen datenbasierter Wertschöpfungspotenziale ist eine konsistente und widerspruchsfreie Datenbasis. Mit dem Forschungsprojekt „Anwendung der Datenfusion bei der Erfassung und Speicherung betrieblicher Rückmeldedaten (DaFuER)“ wird demgemäß das Ziel verfolgt, die Erhöhung der Datenqualität von betrieblichen Rückmeldedaten durch die Anwendung von Methoden der Datenfusion zu ermöglichen.

Als Ergebnis des Forschungsprojekts wird in diesem Leitfaden eine Methode zur anwendungsfallspezifischen Ableitung geeigneter Methoden der Datenfusion dargelegt. Zunächst erfolgt die Definition des Anwendungsfalls. Dabei wird zur Ermittlung relevanter Informationsbedarfe den Anwendenden der Methodik eine Übersicht bereitgestellt, welche die verschiedenen für die Produktionsplanung und -steuerung benötigten Informationen enthält. Außerdem werden Datenquellen anhand der Art der Datenerfassung klassifiziert. Diese Klassifikation ist die Grundlage für die Identifikation der im jeweiligen Anwendungsfall zur Verfügung stehenden Datenquellen.

Im Folgenden werden aus den verfügbaren Datenquellen diejenigen ermittelt, welche fusioniert werden sollen. Dazu wurde eine tabellarische Übersicht erstellt, mit Hilfe derer Datenquellen den Informationen zugeordnet werden, die sie bereitstellen. Weiterhin werden diese Datenquellen hinsichtlich ihrer Datenqualität auf Basis ausgewählter Qualitätsmerkmale bewertet. Für eine benötigte Information wählen die Anwendenden aus den ihnen zur Verfügung stehenden Datenquellen diejenigen zur Fusion aus, welche den Informationsbedarf decken und sich hinsichtlich der Erfüllung der Qualitätsmerkmale komplementieren.

Zuletzt wird eine für den konkreten Anwendungsfall geeignete Fusionsmethode der ausgewählten Datenquellen bestimmt. Grundlage dafür ist eine morphologische Untersuchung von Datenquellen. Durch eine Clusteranalyse möglicher Fehlerarten in Abhängigkeit der Kombination von verschiedenen morphologischen Merkmalsausprägungen werden prozesstypische Fehler der Datenfusion abgeleitet. Somit ist man in der Lage, anhand der ausgewählten Datenquellen die spezifischen Herausforderungen bei der Datenfusion zu identifizieren. Für die finale Auswahl einer für den Anwendungsfall geeigneten Datenfusionsmethode wurden für die ermittelten Prozessfehler die jeweiligen Eignungen der verschiedenen Methoden bewertet. Auf Grundlage dieser Bewertung wählen die Anwendenden schlussendlich diejenige Methode aus, die für die von ihnen identifizierten Herausforderungen am besten geeignet ist.

# 1 Das Forschungsprojekt ‚DaFuER'

In Kapitel eins dieses Leitfadens erhalten die Anwendenden einen groben Überblick über das Ziel des Forschungsprojekts ‚DaFuER' und den daraus entwickelten Leitfaden.

## 1.1 Ziel des Forschungsprojekts ‚DaFuER'

Ziel des Forschungsprojekts „Anwendung der Datenfusion bei der Erfassung und Speicherung betrieblicher Rückmeldedaten" (gen. DaFuER) ist die Sicherstellung der Datenqualität insbesondere für die Produktionssteuerung und das Produktionscontrolling durch die Anwendung der Methoden der Datenfusion und Decision-Fusion auf betriebliche Rückmeldedaten. Durch die mathematische Verknüpfung und den Vergleich von unterschiedlichen Eingangswerten können fehlerhafte Werte identifiziert und somit die Datenqualität signifikant verbessert werden. Dieses Prinzip wird unter dem Begriff der Datenfusion bereits seit längerem für anderen Themenstellungen angewendet.

Das Forschungsprojekt fußt auf der Prämisse, dass die Bedeutung datengetriebener Entscheidungsprozesse im Zuge der Entwicklungen um die digitale Transformation von Unternehmen zunehmend größer wird. Jedoch schafft erst ein möglichst vollständiges, fehlerfreies und echtzeitnahes Abbild (gen. digitaler Schatten) das notwendige Vertrauen der Mitarbeitenden in die vorliegenden Daten und erhöht damit die Akzeptanz der abgeleiteten Entscheidungen. Auf Basis dieser Anforderungen wurden im Projekt drei zentrale Ergebnisse angestrebt:

1. Definition von Kriterien zur Qualität betrieblicher Rückmeldedaten
2. Übertragung von Methoden der Datenfusion auf die Erfassung von Rückmeldedaten produzierender Unternehmen
3. Zuordnung von Anwendungsfällen und Methoden in Form einer Auswahlhilfe

## 1.2 Ziel des Leitfadens

Die Idee des Leitfadens ist es, die Forschungsergebnisse, insbesondere Prozesse, gesammelte Datenquellen und Methoden der Datenfusion, zusammenfassend zu beschreiben und eine Auswahlhilfe für Anwenderunternehmen sowie für Entwickler*innen betrieblicher Anwendungssysteme zu bieten. Hinsichtlich der Anwenderunternehmen lag ein besonderer Fokus auf der Schaffung von Nutzenvorteilen für kleine und mittlere Unternehmen (KMU), weshalb diese bereits frühzeitig in das Projekt integriert wurden. Folgende Unternehmen sind an diesem Forschungsprojekt beteiligt gewesen:

- AUTO HEINEN GmbH, Bad Münstereifel
- Berghof Systeme e. K., Königsee
- DFA Demonstrationsfabrik Aachen GmbH, Aachen
- INDUTRAX GmbH, Hilden
- Maschinenfabrik Möllers GmbH, Beckum
- Mattern Consult Gesellschaft für die Produktionsregelung und Logistik mbH, Ense
- mk Plast GmbH & Co. KG, Monschau
- NETRONIC Software GmbH, Aachen
- SICK AG, Waldkirch
- Ubisense AG, Düsseldorf
- Westaflexwerk GmbH, Gütersloh

Konkret ergibt sich für KMU durch das Projekt im Bereich der operativen Auftragsdurchführung eine Steigerung der Informationstransparenz, was durch die Definition von Qualitätsanforderungen erreicht wird. Zudem wird durch Anwendung der entwickelten Lösungen eine Steigerung der Datenqualität erreicht. Hierdurch werden Entscheidungsprozesse in der Produktionssteuerung und im Produktionscontrolling, aber auch zur Kennzahlenerhebung vereinfacht.

# 2 Grundlagen der Datenfusion

Im zweiten Kapitel werden die Anwendenden hinsichtlich der Notwendigkeit von Datenfusion bei der Erfassung und Speicherung betrieblicher Rückmeldedaten sensibilisiert. Im Anschluss wird ein erster Eindruck über Methoden der Datenfusion und ihre Einsatzmöglichkeiten gegeben.

## 2.1 Problemstellung im Kontext der betrieblichen Rückmeldung

Zentrale Herausforderung der Globalisierung und Digitalisierung für produzierende Unternehmen in Deutschland sind steigende Kundenanforderungen nach individuelleren Produkten bei gleichzeitig stetig kürzer werdenden Lieferzeiten. Insbesondere für kleine und mittlere Unternehmen steigt der Kosten- und Innovationsdruck durch wachsende Konkurrenz aus Niedriglohnländern und einen schrumpfenden Technologievorsprung[1]. Aus diesen Herausforderungen entsteht die Anforderung an eine echtzeitfähige und effiziente Produktionsplanung und -steuerung, die eine fundierte und kurzfristige Entscheidungsfindung erlaubt[2]. Die Grundlage für eine leistungsstarke Produktionsplanung und -steuerung ist eine hohe Informationsverfügbarkeit[3]. Das allein ist jedoch keine hinreichende Bedingung für eine effiziente Produktionsplanung und -steuerung, da nur bei einer ausreichenden Datenqualität eine zuverlässige Entscheidungsfindung möglich ist[4]. Weiterhin bieten die zunehmende Digitalisierung und Vernetzung ein steigendes Nutzenpotenzial für datenbasierte Wertschöpfung, was die Relevanz einer hohen Datenqualität zusätzlich verstärkt[5]. Zentrale Herausforderung der Steigerung der Datenqualität sind die Investitionskosten der Implementierung von entsprechenden Maßnahmen. Dabei zeichnen sich kleine und mittlere Unternehmen insbesondere durch eine eingeschränkte Investitionsfähigkeit aus.[6]

Auf die Erhöhung der Datenqualität eines aggregierten Datensatzes zielt die Datenfusion durch Kombination verschiedener Datenquellen ab. Die Methoden der Datenfusion werden bereits erfolgreich außerhalb der Produktion angewendet, sodass eine Übertragung der Technologien großes Potenzial aufweist[7].

## 2.2 Datenfusion und Datenintegration

Aus der steigenden Menge verfügbarer Daten entstehen erhebliche Nutzenpotenziale. Dies gilt insbesondere für die Anwendung von Verfahren des Data-Minings zur Extraktion von wertschöpfenden Erkenntnissen aus einer Datenmenge. Der Begriff Data-Mining bezeichnet im Kontext der betrieblichen Rückmeldung die nicht triviale Gewinnung von Informationen und Wissen aus Daten und die daraus resultierende Mustererkennung[8].

Es ist naheliegend, dass die Gesamtheit der für die Verfahren des Data-Minings benötigten Daten in den meisten Fällen nicht aus derselben Datenquelle stammt. Weiterhin werden die verwendeten Daten nicht unabhängig voneinander, sondern ganzheitlich betrachtet. Deshalb ist zunächst die Zusammenführung der einzelnen Datenmengen aus den verschiedenen relevanten Datenquellen zu einer gemeinsamen, vollständigen Datenbasis erforderlich. Eine solche Zusammenführung verteilter Daten wird als Datenintegration bezeichnet[9]. Die Datenintegration besteht aus mehreren Schritten, welche in Bild 2.2.1 dargestellt sind. Auf dem Weg von der Erfassung bis zur Anwendung der Daten werden diese zunächst aus verschiedenen Datenquellen extrahiert. Im nächsten Schritt werden die Daten der einzelnen Datenquellen unabhängig voneinander vorverarbeitet. Die drei folgenden Schritte Schema-Matching, Dublettenerkennung und Datenfusion bilden zusammen die Datenintegration.

---

[1] s. Bley et al. 2019, S. 18
[2] s. Schuh et al. 2017, S. 140
[3] s. Nyhuis et al. 2017, S. 33 f.
[4] s. Schuh et al. 2015, S. 200 f.
[5] s. Becker et al. 2017
[6] s. Dienes et al. 2018, S. 46
[7] s. Ruser u. Puente León 2007, S. 94
[8] s. Runkler 2010, S. 24
[9] s. Bleiholder u. Schmid 2015, S. 139

Schließlich werden die aggregierten Datensätze in einen gemeinsamen Bestand exportiert und dem Anwendenden zur Verfügung gestellt. Im Folgenden werden die drei Schritte der Datenintegration näher erläutert (siehe Bild 1).

Der erste Schritt der Datenintegration ist das **Schema-Matching** zur Überwindung von Unterschieden in der Gliederung und dem Aufbau eines Datensatzes, zum Beispiel in Form von verschiedenen Datenmodellen oder Schemata. Dabei wird eine Abbildung erstellt, welche jedem Attribut einer Datenquelle die semantisch äquivalenten Attribute einer anderen Datenquelle zuordnet[10]. Das Ergebnis des Schema-Matchings ist ein Mapping, das diejenigen Attribute mehrerer Datensätze aufeinander abbildet, die inhaltlich den gleichen Aspekt eines Objekts oder Sachverhalts beschreiben. So wird möglicherweise das Attribut „Anschrift" in einer Tabelle dem Attribut „Adresse" in einer anderen Tabelle zugeordnet.

Der zweite Schritt der Datenintegration ist die **Dublettenerkennung**. Dubletten sind definiert als zwei Datensätze, die dasselbe Realweltobjekt repräsentieren[11]. Typische Beispiele für Dubletten sind mehrfach geführte Kunden oder doppelt gebuchte Bestellungen. Bei der Durchführung der Dublettenerkennung werden alle Datensätze paarweise miteinander verglichen und für jedes Paar einer Kennzahl durch die Anwendung eines vorher definierten Ähnlichkeitsmaßes ermittelt. Wenn die ermittelte Kennzahl für ein Paar eine bestimmten Schwellenwert übersteigt, werden die entsprechenden Datensätze als Dubletten gekennzeichnet[12]. So erhält beispielsweise jede Zeile einer Tabelle eine bestimmte Identifikationsnummer, die angibt, welches Realweltobjekt diese Zeile repräsentiert. Zwei Zeilen, die als Dubletten erkannt werden, enthalten dementsprechend dieselbe Identifikationsnummer[13].

Der letzte Prozessschritt der Datenintegration ist die Datenfusion. Ziel der **Datenfusion** ist die Aggregation der erkannten Dubletten, sodass im Ergebnis pro erfasstes Realweltobjekt nur noch ein widerspruchslos repräsentierender Datensatz existiert. Mehrere Repräsentationen desselben Realweltobjekts werden also zu einer einzigen Repräsentation zusammengefasst[14].

In dem Schritt der Datenfusion können die einzelnen Datensätze einer Dublette in verschiedenen Konfliktverhältnissen zueinander stehen[15]:

- Gleichheit: Gleichheit zweier Datensätze bedeutet, dass deren Werte für alle Attribute vollständig übereinstimmen.
- Subsumption: Ein Datensatz subsumiert einen anderen, wenn er weniger Nullwerte als der andere Datensatz besitzt und die restlichen Werte übereinstimmen.
- Komplementierung: Ein Datensatz komplementiert einen anderen, wenn beide Datensätze sich gegenseitig nicht subsumieren und ein Datensatz

[10] s. Bleiholder u. Schmid 2015, S. 124 f.
[11] s. Farkisch 2011, S. 329
[12] s. Farkisch 2011, S. 330
[13] s. Bleiholder u. Schmid 2015, S. 129
[14] s. Bleiholder u. Schmid 2015, S. 133 f.
[15] s. Leser u. Naumann 2007, S. 344 f.

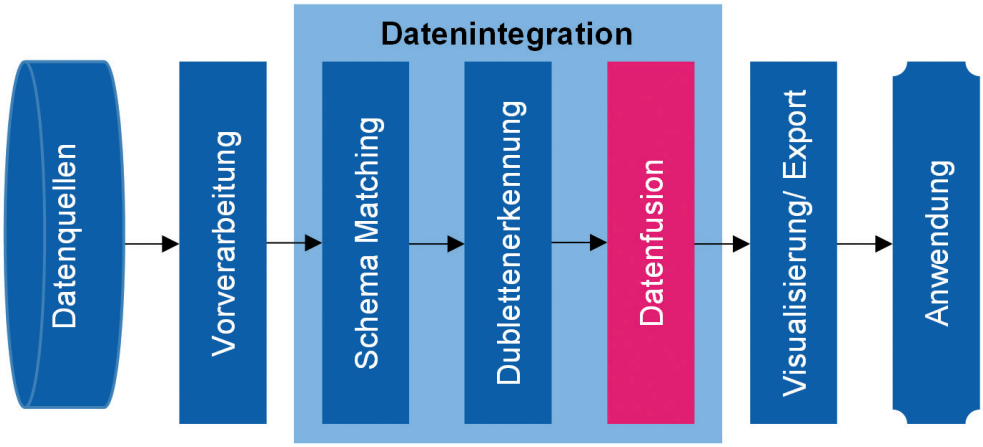


*Bild 1: Prozess der Datenintegration (eigene Darstellung i. A. a. Bleiholder u. Schmid 2015, S. 123)*

für jedes Attribut mit einem Nicht-Nullwert entweder den gleichen Wert wie der andere Datensatz oder der andere Datensatz an dieser Stelle einen Nullwert besitzt.
- Konflikt: Zwei Datensätze können zudem in Konflikt stehen, wenn sie für dasselbe Attribut zwei unterschiedliche Nicht-Nullwerte besitzen.

Der Umgang mit Datenkonflikten bei der Aggregation von Dubletten stellt die zentrale Herausforderung der Datenfusion dar. Gemäß Bild 2 können verschiedene Methoden der Datenfusion zur Vereinigung von in Konflikt stehenden Dubletten angewendet werden. Diese lassen sich in die folgenden Klassen einteilen[16]:

- Konflikte vermeiden: Bei der Vermeidung von Konflikten wird keine Auswahl zwischen den konfligierenden Attributwerten getroffen. Somit werden alle miteinander in Konflikt stehenden Werte übernommen. Daher liegt es in der Hand der Anwendenden, sich nachträglich für einen der Attributwerte zu entscheiden.
- Konflikte ignorieren: Bei der Ignoranz von Konflikten werden Entscheidungen anhand von Entscheidungsregeln getroffen, die die eigentliche Ausprägung der Attribute und die konkreten Datenkonflikte nicht berücksichtigen. Diese Entscheidungsregeln werden einheitlich und unabhängig von der Ausprägung der einzelnen Attributwerte auf die verschiedenen Datensätze angewendet. Dabei wird zwischen instanzbasierten und metadatenbasierten Entscheidungsregeln unterschieden. Letztere treffen Entscheidungen über die Auswahl eines Datensatzes anhand von Metadaten, wie z. B. anhand der Herkunft oder Aktualität einer Datenquelle.
- Konflikte auflösen: Bei der Auflösung von Konflikten werden alle beteiligten Daten betrachtet und anhand von Entscheidungsregeln in Abhängigkeit des konkreten Datenkonflikts bestimmte Attributwerte ausgewählt. Dabei wird erneut zwischen instanz- und metadatenbasierten Entscheidungsregeln unterschieden. Außerdem wird zwischen entscheidenden und vermittelnden Strategien differenziert. Während entscheidende Strategien einen der vorhandenen Attributwerte für die Übernahme in den finalen Datensatz auswählen, ist die Wahl eines Attributwertes, der nicht in den zu integrierenden Datensätzen existiert, bei vermittelnden Strategien ebenfalls möglich.

[16] S. Bleiholder u. Schmid 2015, S. 135 ff.

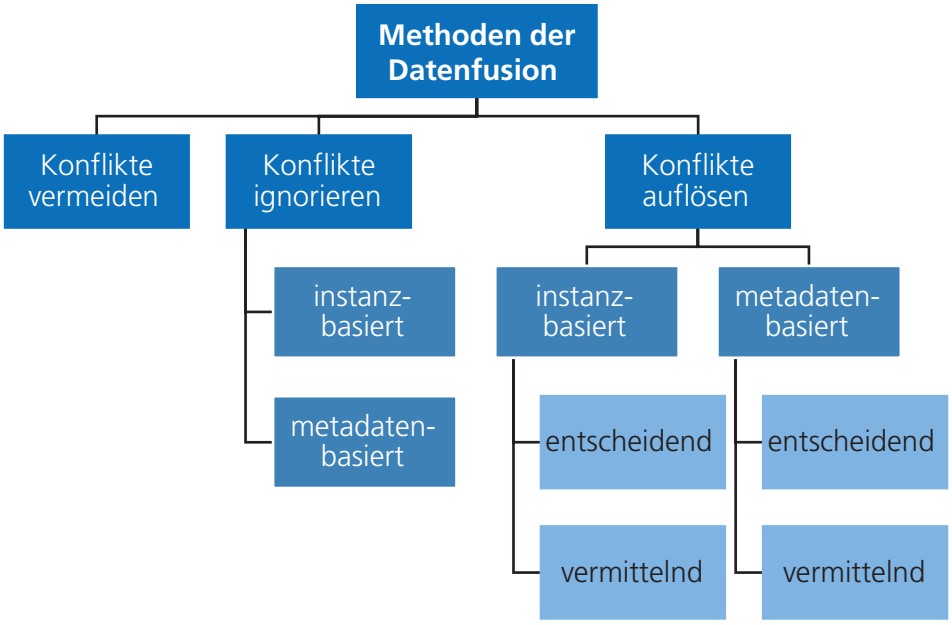


*Bild 2: Modell zur Klassifizierung der Methoden der Datenfusion (eigene Darstellung i. A. a. Bleiholder u. Naumann 2008, S. 8)*

Neben der Zusammenführung einzelner Datensätze im Rahmen der Datenfusion können weiterhin Informationen und Entscheidungen zusammengeführt werden. Dieses Konzept wird als *Decision Fusion* (dt. Entscheidungsfusion) bezeichnet[17]. Grundsätzlich entspricht das Prinzip der Entscheidungsfusion dem der Datenfusion, nur sind anstelle von Datensätzen mehrere möglicherweise konfligierende Entscheidungen über denselben Sachverhalt zu einer Gesamtentscheidung zu aggregieren. Die Strategien der Datenfusion zur Behandlung der Konflikte lassen sich dabei analog auf die Entscheidungsfusion anwenden.

Die Darstellung der verschiedenen Methoden zeigt die Vielfalt, wie mit Konfliktverhältnissen in den jeweiligen Datensätzen umgegangen werden kann. Die Auswahl der geeigneten Methode ist dabei stark von dem Anwendungsfall wie auch den entsprechenden Datensätzen abhängig. Für die Praxis in der betrieblichen Rückmeldung bedeutet das: Die Auswahl einer konkreten Datenfusionsstrategie und deren Algorithmus ist abhängig von der Informationsverfügbarkeit der Rückmeldedaten. Denn in Abhängigkeit der verfügbaren Informationen innerhalb einer Dublette entstehen die oben genannten Konfliktverhältnisse, sodass die entsprechenden Informationsbedarfe und -verfügbarkeit als Ausgangspunkt der beschriebenen Methodik dienen.

[17] s. Fauvel et al. 2006, S. 1

# 3 Allgemeines Vorgehen zur Anwendung der Datenfusion

Der allgemeine Aufbau des Vorgehens wird in Bild 3 deutlich. In einem ersten Schritt wird der jeweils vorliegende Anwendungsfall definiert. Darauf aufbauend werden die zu fusionierenden Datenquellen bestimmt und in einem finalen Schritt geeignete Methoden der Datenfusion ausgewählt.

Die Definition des Anwendungsfalls beginnt mit der Ermittlung der relevanten Informationsbedarfe und der Informationsverfügbarkeit. Dafür werden zunächst die für die Produktionsplanung und -steuerung relevanten Informationen ermittelt. Dies geschieht, indem die Anwendenden der Methodik aus einer Übersicht diejenigen Informationen auswählen, die für sie durch eine mangelhafte Datenqualität unzuverlässig sind (subjektiver Eindruck im operativen Geschäft). Die Menge dieser Informationen bildet den für den Anwendungsfall spezifischen Informationsbedarf. Die Informationsverfügbarkeit wird aus der Gesamtheit der verfügbaren Daten erzeugt. Diese ergeben sich aus den dem Unternehmen zur Verfügung stehenden Datenquellen.

Es folgt die Bestimmung der zu fusionierenden Datenquellen. Für jede als relevant identifizierte Information werden in einer Übersicht potenziell zugehörige Datenquellen aufgelistet, aus denen sich diese Informationen extrahieren lassen. Die Anwendenden sind so in der Lage, ihrem Informationsbedarf konkrete, verfügbare Datenquellen zuzuordnen. Weiterhin wird für jede Datenquelle qualitativ ermittelt, in welchem Maß sie verschiedene Kriterien der Datenqualität erfüllt. Auf Basis dieser Qualitätsmerkmale ist es möglich, diejenigen verfügbaren Datenquellen für die Fusion auszuwählen, die sich gegenseitig in Bezug auf die Erfüllung der Datenqualitätsmerkmale komplementieren.

Schließlich erfolgt die Auswahl geeigneter Methoden der Datenfusion. Auf Basis einer Klassifikation der betrachteten Datenquellen wurde eine Morphologie zur generischen Beschreibung einer Datenquelle entwickelt. Durch Kombination aller möglichen Arten von Datenquellen bzw. deren morphologischer Attribute können prozesstypische Fehler bei der Fusion abgeleitet und Fehlerklassen gebildet werden. Bei den prozesstypischen Fehlern wie auch den entsprechenden Fehlerklassen handelt es sich um theoretische Problemstellungen bei der Integration der verschiedenen Datenquellen, welchen durch die Auswahl der Daten-

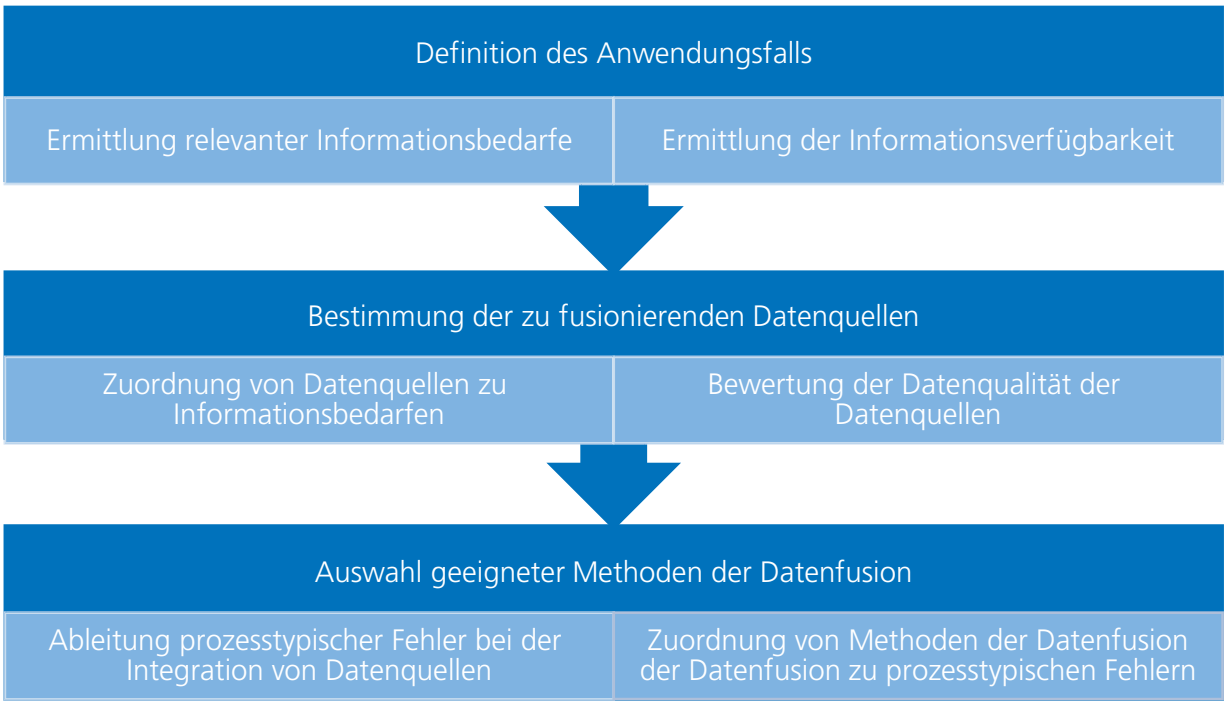


*Bild 3: Aufbau des Leitfadens (eigene Darstellung)*

fusionsmethodik entgegengewirkt wird. Im nächsten Schritt wird für jede der beschriebenen Fehlerklassen ermittelt, inwieweit eine Methode für die Anwendung der Datenfusion resistent gegen den 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, die bestmöglich die zentralen Herausforderungen bei der Kombination der gewählten Datenquellen bewältigen.

Im Rahmen dieses Leitfadens werden den Anwendenden zwei Möglichkeiten zur konkreten Umsetzung des hier dargelegten Konzepts geboten: Eine Möglichkeit ist das schrittweise Durchlaufen und Abarbeiten der einzelnen Prozessschritte dieses Leitfadens, wie sie im Folgenden dargestellt werden. Als Hilfestellung kann dafür die Checkliste zur Anwendung der Datenfusion im Kontext betrieblicher Rückmeldedaten (s. Kapitel 9, S. 55) dienen. Eine Alternative bietet das im Rahmen dieses Forschungsprojekts erstellte interaktive Online-Tool, in welchem die Anwendenden Schritt für Schritt durch das oben dargestellte Vorgehen zur Durchführung der Datenfusion angeleitet und geführt werden.

Das Online-Tool finden Sie hier:

dafuer-tool.fir.de

# 4 Detailvorgehen bei der Anwendung der Datenfusion

Das zuvor erläuterte allgemeine Vorgehen zur Anwendung der Datenfusion wird in diesem Kapitel durch die detaillierte Darstellung der einzelnen Arbeitsschritte konkretisiert. Dazu wird in einem ersten Schritt ein Verfahren zur Definition des Anwendungsfalls dargelegt. Darauf aufbauend folgt die Bestimmung der zu fusionierenden Datenquellen. Der finale Schritt – und Ziel dieses Leitfadens – ist die Ableitung geeigneter Methoden der Datenfusion zur konkreten Anwendung in der Praxis.

## 4.1 Definition des Anwendungsfalls

Im Folgenden wird ein Verfahren zur Definition des vorliegenden Anwendungsfalls dargelegt. Dazu wird im ersten Hauptschritt eine Strategie zur Ermittlung des spezifischen Informationsbedarfs entworfen und untersucht, wie die vorliegende Informationsverfügbarkeit ermittelt werden kann.

### 4.1.1 Ermittlung relevanter Informationsbedarfe

Im Rahmen des Forschungsprojekts ‚DaFuER' werden Informationsbedarfe als Summe aller Informationen bezeichnet, welche für die Bearbeitung der einzelnen Prozesse der Produktionsplanung und -steuerung benötigt werden. Hinsichtlich der Informationsbedarfe bzw. benötigten Daten kann zwischen Stamm- und Bewegungsdaten unterschieden werden.

- Stammdaten: Stammdaten existieren über einen längeren Zeitraum und besitzen eine geringe Änderungshäufigkeit. Sie existieren unabhängig von spezifischen Buchungen und stellen den Rahmen für die Planung und Kontrolle von Transaktionen dar. Somit sind Stammdaten eigenschaftsorientierte Daten, die existenziell unabhängig von anderen betrieblichen Datenarten sind und der Identifizierung von Kernelementen sowie Geschäftsobjekten dienen. Die relevantesten Stammdaten sind Materialdaten, Stücklisten, Maschinendaten, Ressourcen und Arbeitspläne[18].
- Bewegungsdaten: Bewegungsdaten sind dynamische Daten, die im Kontext der Datenverwaltung die Informationen darstellen, welche aus Transaktionen erfasst werden. Bewegungsdaten können logistischer oder arbeitsbezogener Natur sein und von einer Bestellung über den Versandstatus bis hin zu den geleisteten Arbeitsstunden reichen. Typische Bewegungsdaten sind zum Beispiel Lagerbestands- und Produktionsdaten[19].

Ein Auszug der von den verschiedenen Prozessen der Produktionsplanung und -steuerung (PPS) benötigten Informationen ist in Bild 4 mit dem Fokus auf Bewegungsdaten dargestellt.

[18] s. Loos 1999, S. 2 f.; s. Kurbel 2013, S. 20

[19] s. Loos 1999, S. 2; s. Kurbel 2013, S. 20; s. Schuh et al. 2012, S. 78

| | | | Losgrößenrechnung | Feinterminierung | Ressourcenfeinplanung | Reihenfolgeplanung | Verfügbarkeitsprüfung | Auftragsfreigabe |
|---|---|---|---|---|---|---|---|---|
| Stammdaten | Materialstammdaten | Materialnummer | x | x | x | x | x | x |
| | | Lagerinformationen | x | x | | | x | |
| | | Kosteninformationen | x | | | x | | |
| | Produktionsdaten | Losgröße | x | | | | | |
| | | Standard-Plan-Durchlaufzeit | x | x | x | | | |
| | | Standard-Arbeitsplan-Nummer | x | x | x | x | x | x |
| | | Arbeitsgangnummer | x | | | x | x | x |
| | | Gesamtbedarfsmenge | | | | x | x | x |
| | Arbeitsplandaten | Arbeitsplannummer | x | x | x | x | x | x |
| | | Arbeitsplan-Variantennummer | x | x | x | x | x | x |
| | | Arbeitsplatz | | | x | x | x | x |
| | | Zeitdauer Durchführung | x | x | x | x | x | x |
| | | benötigte Fertigungshilfsmittel | x | x | x | x | x | x |
| | | Belegungszeitfaktor | | x | x | | | |

*Bild 4: Informationsbedarfe der Produktionsplanung und -steuerung (Teilauszug bzgl. der Bewegungsdaten) (eigene Darstellung)*

Eine vollständige Übersicht der Informationsbedarfe eines PPS-Systems, aufgeschlüsselt nach Bewegungs- und Stammdaten, kann Bild 16 im Anhang (s. S. 27) entnommen werden.

Im Rahmen der betrieblichen Rückmeldung und dieses Leitfadens lässt sich die Produktionsplanung und -steuerung in insgesamt sechs Teilbereiche unterteilen. Dazu gehören die Losgrößenrechnung, die **Feinterminierung**, die **Ressourcen-** und **Reihenfolgeplanung**, die **Verfügbarkeitsprüfung** sowie die **Auftragsfreigabe**. Jeder dieser Teilbereiche ist durch seinen individuellen Informationsbedarf gekennzeichnet, welcher im obigen Bild hinsichtlich Bewegungsdaten aufgeschlüsselt und nochmals in Ressourcen- und Auftragsdaten differenziert wird. Zu den Ressourcendaten zählen im Rahmen der Bewegungsdaten Lagerbestände und Bedarfe. Zu den Auftragsdaten zählen u. a. die Auftragsnummer, Auftrags- und Arbeitsgangzeiten und der Auftragsfortschritt. Für die Ermittlung der relevanten Informationsbedarfe wählt der Anwendende zunächst den oder die Teilbereiche der Produktionsplanung und -steuerung aus, deren Datenqualität durch die Durchführung der Datenfusion optimiert werden soll. Durch die Betrachtung der Zellen, welche mit einem Kreuz markiert sind, wird ersichtlich, welche (Bewegungs-) Daten relevante Informationsbedarfe für den jeweiligen Teilbereich darstellen. So lässt sich beispielsweise erkennen, dass für die Losgrößenrechnung Informationen über die Lagerbestände, Bearbeitungszeiten und Maschinenzeiten benötigt werden. Die Auftragsfreigabe hingegen benötigt Informationen über die Bedarfe, die Auftragsnummer, Auftrags- und Arbeitsgangzeiten sowie die produzierte Menge.

Eine Übersicht kurzer Definitionen der im Rahmen des obigen Bildes verwendeten Begrifflichkeiten kann dem angefügten Glossar unter Kapitel 8.1 (S. 45) entnommen werden.

### 4.1.2 Ermittlung der Informationsverfügbarkeit

Dem Informationsbedarf steht die Informationsverfügbarkeit gegenüber. Diese ergibt sich aus Datenquellen, welche den Anwendenden zur Verfügung stehen. Im Zuge dieses Forschungsprojekts werden Datenquellen in Anlehnung an die *VDI*-Richtlinie 5600 anhand der Art der Datenerfassung klassifiziert[20]. Die Klassifikation von für die PPS relevanten Datenquellen ist in Bild 5 dargestellt. Sie dient als erste Übersicht und Orientierungshilfe zur Identifikation der zur Verfügung stehenden Datenquellen, aus welchen dann verfügbare Informationen strukturiert abgeleitet werden können.

[20] s. VDI 2016, S. 41

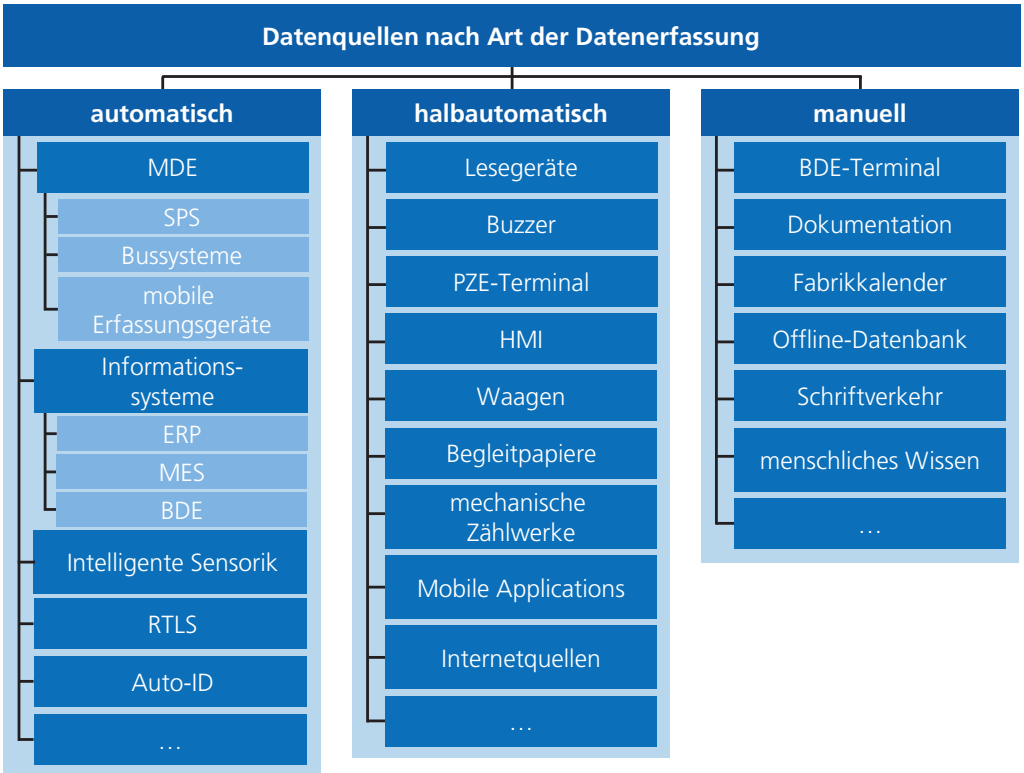


*Bild 5: Klassifikation von Datenquellen nach Art der Erfassung (eigene Darstellung)*

Bei der Betrachtung der Datenquellen wird zwischen der automatischen, halbautomatischen und manuellen Datenerfassung unterschieden. Die automatische Datenerfassung erfordert keine Bedieneingriffe und erfolgt zyklisch oder durch die Auslösung eines bestimmten Ereignisses. Die halbautomatische Datenerfassung erfordert das manuelle Starten der Erfassung, zeichnet die zu erfassenden Werte jedoch automatisch auf. Die manuelle Datenerfassung erfolgt durch eine händische Eingabe mithilfe eines Eingabegeräts in ein Eingabefeld oder ein für die Erfassung vorgesehenes Formular. Ein Beispiel für eine Datenquelle mit automatischer Datenerfassung sind Informationssysteme, welche sich in die Bereiche ERP, MES und BDE unterteilen lassen. Im Allgemeinen sind Informationssysteme als Drehscheibe des unternehmensinternen Datentransfers zu verstehen, in dem relevante Daten zentral erfasst und verarbeitet werden. Ein ERP(Enterprise-Resource-Planning)-System dient der Unterstützung sowie der Bündelung und Steuerung aller notwendigen Geschäftsprozesse innerhalb eines Unternehmens.

Weitere Informationen zu den einzelnen Datenquellen können dem Glossar unter Kapitel 8.2 entnommen werden.

## 4.2 Bestimmung der zu fusionierenden Datenquellen

Im nächsten Hauptschritt wird ein Verfahren zur Bestimmung der zu fusionierenden Datenquellen dargelegt. Dafür wurde eine Übersicht zur Zuordnung von Datenquellen zu spezifischen Informationsbedarfen entwickelt. In einem zweiten Schritt kann dann die Datenqualität der jeweiligen Datenquellen für ausgewählte Qualitätsmerkmale bewertet werden.

### 4.2.1 Zuordnung von Datenquellen zu Informationsbedarfen

Zunächst ordnen die Anwendenden ihren spezifischen Informationsbedarfen Datenquellen zu, die diesen Informationsbedarf decken. Eine allgemeine Zuordnung der in Bild 5 (s. S. 14) dargestellten Datenquellen zu denjenigen Informationsbedarfen, die sie potenziell erfüllen, ist auszugsweise für Stammdaten in Bild 6 gegeben. Eine äquivalente Zuordnung bezüglich Bewegungsdaten kann Bild 17 im Anhang (s. S. 28) entnommen werden.

| | | Informationsbedarf | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Materialnummer | Lagerinformationen | Kosteninformationen | Losgröße | Standard-Plan-Durchlaufzeit | Standard-Arbeitsplan-Nummer | Arbeitsgangnummer | Gesamtbedarfsmenge | Arbeitsplannummer | Arbeitsplan-Variantennummer | Arbeitsplatz | Zeitdauer Durchführung | benötigte Fertigungshilfsmittel | Belegungszeitfaktor | Maschinenkapazität | Instandhaltungsdaten |
| Informationsverfügbarkeit | MDE | | | | | | | x | | | | x | x | x | | x | x |
| | Informationssysteme | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x |
| | Intelligente Sensorik | | | | | | | | | | | | x | | | | |
| | RTLS | | x | | | | | | x | | | x | | | x | | |
| | Auto-ID | x | x | x | | | | | x | | | | | | | | |
| | Lesegeräte | x | | | | | | | | | | x | | | | | |
| | Buzzer | | | | | | | | | | | | | | | | x |
| | PZE-Terminal | | | | | | | | | | | x | x | x | x | | |
| | HMI | x | | | | | | x | | | | x | x | x | x | x | x |
| | Waagen | | x | | | | | | | | | | | | | | |
| | Begleitpapiere | x | x | x | x | x | x | x | x | x | x | | | | | | |
| | mechanische Zählwerke | | | | | | | | | | | | | | | | |
| | Mobile Applications | x | x | x | | x | x | | x | x | x | | x | x | x | x | |
| | Internetquellen | x | x | x | | | | | | | | | | x | | x | |
| | BDE-Terminal | x | x | x | | | | | | x | | x | x | | x | | x |
| | Dokumentation | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x |
| | Fabrikkalender | | | | | x | x | | | x | x | | x | | | | |
| | Offline-Datenbank | x | x | x | x | x | x | | x | | x | x | | x | x | x | |
| | Schriftverkehr | x | x | x | x | x | x | | x | | x | x | | x | | | |
| | menschliches Wissen | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x |

*Bild 6: Zuordnung von Datenquellen zu Informationsbedarfen (Stammdaten) (eigene Darstellung)*

Die Informationsbedarfe der Produktionsplanung und -steuerung, aufgeschlüsselt nach Stammdaten, wie z. B. der Materialnummer, Lagerinformationen oder Kosteninformationen, können spaltenweise der horizontalen Achse im oberen Teil des Bildes entnommen werden. Auf der vertikalen Achse sind die oben genannten Datenquellen gelistet, welche potenziell die benötigten (Stamm-)Daten bereitstellen können. Durch die Betrachtung der Zellen, welche mit einem Kreuz markiert sind, wird ersichtlich, welche Informationsbedarfe durch eine Datenquelle erfüllt werden können. Die Maschinendatenerfassung (MDE) beispielsweise ist ein System zur automatisierten Dokumentation von im Rahmen des Produktionsprozesses unmittelbar an Maschinen und Anlagen entstehenden Informationen. Hinsichtlich der Betrachtung von Stammdaten können diese Informationen Aufschluss geben über die Arbeitsgangnummer, den Arbeitsplatz, die Zeitdauer der Durchführung, benötigte Fertigungshilfsmittel, die Maschinenkapazität und Instandhaltungsdaten.

Es gilt zu berücksichtigen, dass sich die dargestellten Zuordnungen in der Übersicht nur auf eine potenzielle Deckung des Informationsbedarfs beziehen. So lassen sich z. B. nicht aus jeder Maschinensteuerung Daten bezüglich der Instandhaltung der Maschine ableiten.

### 4.2.2 Bewertung der Datenqualität der Daten

Im zweiten Schritt der Bestimmung der zu fusionierenden Datenquellen wird die Datenqualität der einzelnen Datenquellen bewertet und miteinander verglichen. Im ersten Schritt haben die Anwendenden anhand der Zuordnung von Datenquellen zu Informationsbedarfen (s. Bild 6, S. 15) eine Vorauswahl von relevanten Datenquellen getroffen. Aus dieser Vorauswahl werden nun diejenigen Datenquellen zur Fusion ausgewählt, welche sich bezüglich ihrer Ausprägung in verschiedenen Qualitätsmerkmalen komplementieren. Insbesondere Wang u. Strong haben mit ihrem empirisch erhobenen Qualitätsmodell eine verbreitete Grundlage zur Einteilung von Qualitätsmerkmalen gelegt. Auf diesen Ergebnissen aufbauend wurde von der *DGIQ (Deutsche Gesellschaft für Informations- und Datenqualität)* ein Modell für den deutschsprachigen Raum erarbeitet. Dieses Modell betrachtet fünfzehn Dimensionen der Informationsqualität. Wie in Bild 7 ersichtlich, können diese Dimensionen in die Kategorien **System**, **Inhalt**, **Darstellung** und **Nutzung** eingeordnet werden. Eine kurze Erläuterung aller in Bild 7 dargestellten Qualitätsmerkmale kann dem Glossar unter Kapitel 8.3 entnommen werden. Im Rahmen dieses Leitfadens wurde sich jedoch auf die folgenden in der Praxis relevantesten Qualitätsmerkmale beschränkt[21]:

- Vollständigkeit: Daten werden als vollständig bezeichnet, wenn zu einem festgelegten Zeitpunkt alle für einen Prozessschritt benötigten Daten zur Verfügung stehen.
- Aktualität: Die Aktualität eines Datensatzes beschreibt die Fähigkeit, bei Änderungen in der realen Welt zeitnah die entsprechenden Daten anzupassen. Aktualität ist somit die Resistenz eines Datensatzes gegenüber Fehlern aufgrund der zeitlichen Änderung der realen Welt.

[21] s. Rohweder et al. 2015, S. 30 – 38; s. Apel et al. 2009, S. 22 f.

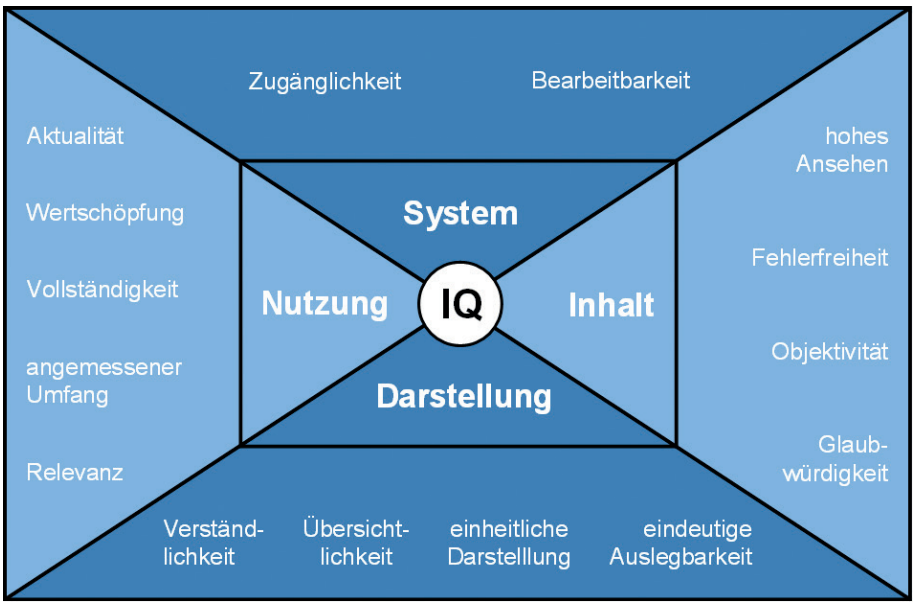


*Bild 7: Qualitätsmodell der DGIQ (eigene Darstellung i. A. a. Rohweder et al. 2015, S. 30)*

- Fehlerfreiheit: Daten werden als fehlerfrei bezeichnet, wenn sie mit der Realität widerspruchsfrei übereinstimmen.
- Zugänglichkeit: Daten werden als zugänglich bezeichnet, sofern diese anhand von einfachen Verfahren und auf direktem Weg abrufbar sind.
- Objektivität: Daten werden als objektiv bezeichnet, sofern sie sachlich und wertfrei, also ohne subjektiven Einfluss sind.
- Genauigkeit: Daten werden als genau bezeichnet, sofern sie in Abhängigkeit des jeweiligen Anwendungsfalls als korrekt und zuverlässig angesehen werden können. Im Hinblick auf Bild 7 (s. S. 16) ist die Genauigkeit inhaltlich als eine Kombination der Qualitätsmerkmale **Relevanz** und **Vollständigkeit** zu interpretieren.

Die übergreifende Bewertung Datenquellen (Auszug) auf Basis der genannten Qualitätsmerkmale ist im Folgenden in Bild 8 ersichtlich. Eine vollständige Bewertung aller in Bild 5 (s. S. 14) gelisteten Datenquellen kann Bild 18 im Anhang (s. S. 29) entnommen werden.

Auf der horizontalen Achse sind die obigen Datenqualitätsmerkmale aufgeführt, die vertikale Achse listet einen kurzen Auszug der im Rahmen dieses Leitfadens betrachteten Datenquellen auf. Ein vollständig rot ausgefüllter Kreis z. B. in der Spalte „Aktualität" bedeutet, dass die Datenquelle eine sehr hohe Aktualität besitzt. Umgekehrt bedeutet ein vollständig grau ausgefüllter Kreis in der Spalte „Aktualität", dass die Datenquelle eine sehr geringe Aktualität besitzt. Die einzelnen Bewertungen entsprechen dabei grundsätzlich allgemeinen Tendenzen und keinen endgültigen Dogmen. Deshalb sollten jedoch die hier durchgeführte allgemeine Betrachtung und Bewertung der Datenqualität für jeden konkreten Anwendungsfall anhand der erarbeiteten Grundlagen individuell erfolgen.

Die MDE beispielsweise besitzt eine hohe Vollständigkeit, da oftmals Informationen über den produzierten Ausschuss und ggf. auch über dessen Ursache abgeleitet werden können. Abhängig von der tatsächlichen Anlage besitzt die MDE mit einer Zykluszeit von unter einer Sekunde eine sehr hohe Aktualität. Maschinendaten sind darüber hinaus robust gegen stochastische Fehler, unterliegen aber teilweise regelmäßigen Fehlern, wie beispielsweise einer fehlerhaften Implementierung einer Erfassungsschnittstelle. Der mit solch einer Implementierung verbundene Programmieraufwand ist der Grund für eine nur geringe Zugänglichkeit. Weiterhin sind maschinell erfasste Daten sehr objektiv und genau.

Zur Optimierung der Zugänglichkeit der Daten, welche aus einer MDE gewonnen werden können, kann diese mit einem Informationssystem, beispielsweise einem ERP-System, fusioniert werden. Denn diese besitzen eine vergleichsweise hohe Zugänglichkeit. Generell gilt: Der Anwendende wählt im Optimalfall diejenigen Datenquellen zur Fusion aus, welche sich bezüglich ihrer Ausprägung in verschiedenen Qualitätsmerkmalen komplementieren.

## 4.3 Auswahl geeigneter Methoden der Datenfusion

Der dritte und letzte Hauptschritt befasst sich mit der finalen Auswahl geeigneter Methoden der Datenfusion. Begonnen wird mit der systematischen Untersuchung möglicher Kombinationen von Datenquellen zur Ableitung möglicher prozesstypischer Fehler bei der Integration. Anschließend werden dann den ermittelten Fehlerarten geeignete Methoden der Datenfusion zugeordnet.

| | Vollständigkeit | Aktualität | Fehlerfreiheit | Zugänglichkeit | Objektivität | Genauigkeit |
|---|---|---|---|---|---|---|
| MDE |  |  |  |  |  |  |
| Informationssysteme |  |  |  |  |  |  |
| Intelligente Sensorik |  |  |  |  |  |  |

*Bild 8: Bewertung der Datenqualität der Datenquellen (Auszug) (eigene Darstellung)*

### 4.3.1 Ableitung prozesstypischer Fehler

Um prozesstypische Fehler ableiten zu können, müssen die Datenquellen im ersten Schritt detailliert beschrieben werden. Dazu bietet sich die in Bild 9 (s. S. 18) dargestellte Morphologie an:

Auf der vertikalen Achse in blau sind die für die morphologische Untersuchung der Datenquellen ausgewählten Merkmale gelistet. Sie können hinsichtlich unterschiedlicher Ausprägungen differenziert werden, welche in der zum Merkmal gehörigen Zeile in grau aufgeführt sind. Konkret beschreiben lassen sich die Merkmale wie folgt beschreiben:

- Herkunft: Die Herkunft beschreibt den Ursprungsort der Datenquellen.
- Metadaten: Metadaten enthalten Informationen über den Aufbau, die Struktur und den Inhalt der betrachteten Daten[22].
- Art der Erfassung: Die Art der Erfassung wird anhand des Automatisierungsgrades differenziert[23].
- Erfassungsauslösung: Die Erfassungsauflösung ist die Ursache der Datenerfassung[23].
- Archivierungszeitraum: Der Archivierungszeitraum ist der Zeitraum, über welchen erfasste Daten gespeichert werden.
- Aktualisierungsrate: Die Aktualisierungsrate ist die Zeitspanne zwischen der Speicherung eines Datensatzes und dessen Aktualisierung.
- Schnittstellen: Schnittstellen stellen eine Möglichkeit dar, die von einer Datenquelle generierten Daten abzugreifen.
- Datentyp: Ein Datentyp bezeichnet eine Menge von Datenobjekten, welche die gleiche Struktur haben und mit denen die gleichen Operationen ausgeführt werden können[25].
- Datenstruktur: Die Datenstruktur ist eine Möglichkeit der Datenspeicherung und Organisation[26].
- Skalenniveau: Das Skalenniveau beschreibt die Skalierbarkeit einer Messung[27].
- Auflösung: Die Auflösung ist der kleinstmögliche Abstand zwischen einem Messwert und dem nächsthöheren[28].
- Störanfälligkeit: Die Störanfälligkeit ist ein Maß für die Robustheit der Datenerfassung gegenüber äußeren Einflüssen und für die Resistenz gegenüber Datenfehlern.

[22] s. Apel et al. 2009, S. 208
[23] s. VDI 2016, S. 42
[24] s. Reinhardt 1996, S. 8 f.
[25] s. Lange u. Stegemann 1985, S. 7
[26] s. Schramm 2008, S. 9 f.
[27] s. Niederée u. Mausfeld 1996, S. 385 f.
[28] s. Weichert u. Wülker 2010, S. 11

| | | | | |
|---|---|---|---|---|
| **Herkunft** | intern | | extern | |
| **Metadaten** | vorhanden | teilweise vorhanden | nicht vorhanden | |
| **Art der Erfassung** | automatisch | halbautomatisch | manuell | |
| **Erfassungs-auslösung** | kontinuierlich | zyklisch | getriggert | |
| **Archivierungs-zeitraum** | > 1 Monat | > 1 Woche | > 1 Tag | < 1 Tag |
| **Aktualisierungs-rate** | > 1 Stunde | < 1 Stunde | < 1 Minute | < 1 Sekunde |
| **Schnittstellen** | Maschinen-schnittstelle | Software-schnittstelle | Mensch-Maschine | Mensch-Mensch |
| **Datentyp** | Integer | Float | Boolean | String |
| **Datenstruktur** | strukturiert | semistrukturiert | unstrukturiert | |
| **Skalenniveau** | Nominal-skala | Ordinal-skala | Intervall-skala | Verhältnis-skala |
| **Auflösung** | > 10 % | < 10 % | < 1 % | < 0,1 % |
| **Störanfälligkeit** | hoch | mittel | gering | |

*Bild 9: Morphologie zur Einteilung von Datenquellen (eigene Darstellung)*

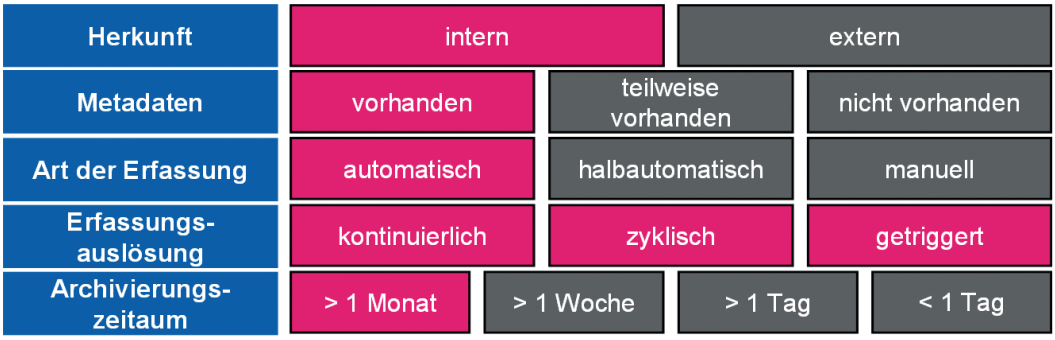


*Bild 10: Morphologie einer speicherprogrammierbaren Steuerung (Auszug) (eigene Darstellung)*

In Bild 10 wird auf die erläuterte Morphologie auszugsweise anhand einer speicherprogrammierbaren Steuerung genauer eingegangen.

Eine speicherprogrammierbare Steuerung gehört als fester Bestandteil der Maschinensteuerung zu den internen Datenquellen. Metadaten sind bei der speicherprogrammierbaren Steuerung zum Beispiel in Form von bestimmten Parametrisierungsübersichten vorhanden. Die Datenerfassung erfolgt automatisch. Eine kontinuierliche, zyklische, sowie getriggerte Datenerfassung ist jeweils implementierbar. Die Eingangs- und Ausgangswerte werden in jedem Zyklus aktualisiert.

Die erläuterte Morphologie wurde exemplarisch für eine gewisse Auswahl von Datenquellen angewendet und kann dem Anhang (s. Bild 19, S. 30; Bild 20, S. 30; Bild 21, S. 31) entnommen werden.

Eine tiefergehende Aufschlüsselung der erstellten Morphologie mit weiterführenden Definitionen und Erklärungen zu den obigen Merkmalen sowie eine vollständige Übersicht der morphologischen Ausprägungen aller im Rahmen dieses Forschungsprojekts betrachteten Datenquellen können den Bildern 22 bis Bild 25 (s. S. 32 – 35) im Anhang sowie dem Glossar unter Kapitel 8.4 (s. S. 49 ff.) entnommen werden.

Durch die Kombination von verschiedenen Datenquellen werden auch verschiedene morphologischen Ausprägungen miteinander verbunden. Dies wiederum kann bei einer Integration zu potenziellen prozesstypischen Fehlern bzw. Fehlerklassen führen. Dies lässt sich exemplarisch in Bild 11 durch die Kombination für das Merkmal der Herkunft nachvollziehen.

In verschiedenen Unternehmen können beispielsweise unterschiedliche Standards bei der Strukturierung und Darstellung von Daten herrschen. Somit liegen bei Datenquellen aus verschiedenen Unternehmen potenziell jeweils unterschiedliche Datenschemata oder eine unterschiedliche Syntax der Datensätze vor. Außerdem sind bei der Verwendung von ausschließlich externen Datenquellen eventuell keine oder nicht genügend Metadaten vorhanden, da ein externes Unternehmen diese üblicherweise nicht zugänglich macht. Zuletzt ist es möglich, dass Datensätze aus unterschiedlichen Unternehmen bei identischer Syntax eine zum Teil unterschiedliche Semantik aufweisen.

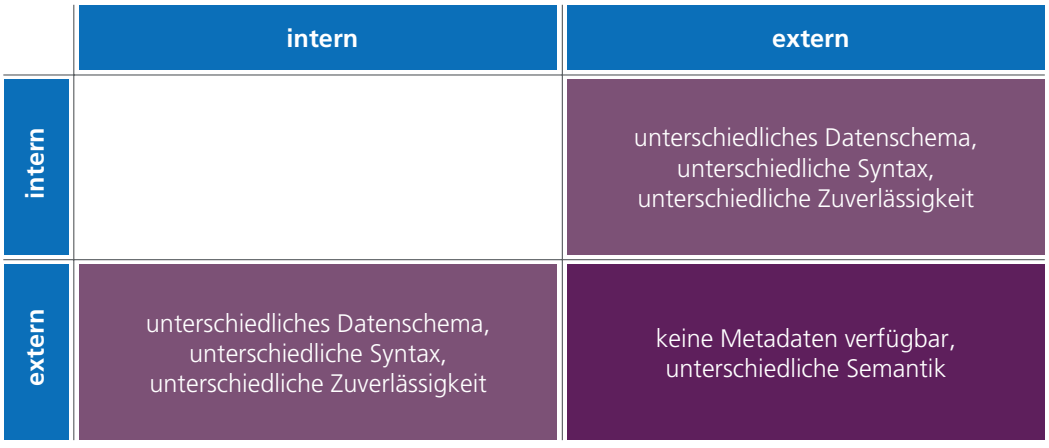


*Bild 11: Ableitung prozesstypischer Fehler – Herkunft (eigene Darstellung)*

Analog zu dem beschriebenen Beispiel wurden für alle Merkmale die prozesstypischen Fehler abgeleitet, welche Bild 26 bis Bild 36 (s. S. 36 – 41) im Anhang entnommen werden können. Daraus lassen sich insgesamt folgende prozesstypische Fehlerkategorien ableiten:

- unterschiedliches Datenschema,
- unterschiedliche Syntax bei gleicher Semantik,
- unterschiedliche Semantik bei gleicher Syntax,
- unterschiedliche Zuverlässigkeit der Datenquellen,
- hohe Varianz innerhalb der Datensätze,
- hohe Subjektivität,
- keine Verfügbarkeit von Metadaten,
- keine Verfügbarkeit von historischen Daten,
- große Datenmengen,
- Erfassung unterschiedlicher Objekte,
- Erfassung unterschiedlich vieler Datensätze pro Objekt,
- unterschiedliche zeitliche Auflösung der Daten,
- geringe diskrete Auflösung der Daten,
- keine mathematischen Operatoren anwendbar,
- fehlende Schlüsselattribute,
- eingeschränktes Schema-Matching,
- unterschiedliches Skalenniveau.

Die Datensätze können sich hinsichtlich des Schemas, der Syntax sowie der Semantik unterscheiden. Weiterhin sind manche Datensätze zuverlässiger als andere. Eine teilweise hohe Subjektivität und Varianz der Datensätze, welche die Behandlung von statistischen Ausreißern notwendig macht, können weitere prozesstypische Fehler sein. Darüber hinaus ist die fehlende Verfügbarkeit von Metadaten und/oder historischen Daten eine der größten Herausforderungen für die Datenfusion. Vor allem bei großen Datenmengen müssen lange Lauf- und Rechenzeiten eingeplant werden. Je nach Kombination von Datenquellen ist es außerdem möglich, dass das gewählte Objekt nur von einer Datenquelle erfasst wurde und nicht in einer anderen Datenquelle repräsentiert ist. Gleichzeitig ist denkbar, dass für ein Objekt eine Datenquelle über eine gewisse Zeitspanne mehr Datensätze erfasst hat als eine andere Quelle. Diese Problematik ist vergleichbar mit dem Fall, bei welchem ein Datensatz eine höhere zeitliche Auflösung besitzt als ein anderer. Eine geringe diskrete Auflösung, also die Auflösung der erfassten Datenwerte selbst, ist bei der Datenfusion ebenfalls potenziell problematisch. Außerdem sind mathematische Operatoren aufgrund der Darstellung oder der Form der Daten zum Teil nicht anwendbar. Fehlende (Schlüssel-) Attribute sowie die eingeschränkte Möglichkeit eines anwendungsgerechten Schema-Matchings sind weitere Herausforderungen der Datenfusion. Schließlich ist ein unterschiedliches Skalenniveau der Daten aufgrund der fehlenden Vergleichbarkeit und der eingeschränkten Anwendungsmöglichkeiten von entsprechenden Operatoren problematisch.

### 4.3.2 Zuordnung von Methoden der Datenfusion zu prozesstypischen Fehlern

Im nächsten Schritt gilt es, verschiedene Methoden der Datenfusion zu sammeln und anhand ihrer Stärken und Schwächen voneinander abzugrenzen. Dadurch können für die jeweiligen Methoden besonders geeignete Anwendungsgebiete aufgezeigt werden. Im Rahmen dieses Forschungsprojekts liegt der Fokus dabei auf den folgenden Datenfusionsmethoden:

- **Entscheidungsregeln**: Sie entsprechen konkreten Heuristiken, die im Falle konfligierender Daten spezifische Handlungsanweisungen geben[30].
- **Klassische Statistik**: Beobachtete, zu fusionierenden Daten werden als empirische Repräsentation einer Zufallsvariablen betrachtet. Die Wahrscheinlichkeitsverteilung dieser Zufallsvariablen ist abhängig von einer dazugehörigen tatsächlichen, jedoch unbekannten Messgröße. Die Schätzung dieser Messgröße anhand der vorliegenden Daten ist das Ergebnis der Datenfusion[30].
- **Bayes'sche Inferenz**: Interpretation der Wahrscheinlichkeiten als Degree-of-Belief (DoB). Der DoB repräsentiert für ein Ereignis den Grad der Überzeugung bezüglich des Eintretens des Ereignisses auf Basis der vorliegenden Daten[31].
- **Dempster-Shafer-Methode**: Erweitert die Wahrscheinlichkeit um das zweidimensionale Maß der Evidenz. Diese setzt sich zusammen aus dem DoB und der Plausibilität, dem Maß für die maximale Möglichkeit der Korrektheit einer Hypothese[32].
- **Fuzzy-Logik**: Modellierung von Ungewissheit oder Vagheit durch eine kontinuierlich abgestufte anstatt absolute Zuordnung von Objekten zu bestimmten Klassen. So wird eine Menge nicht durch die in ihr enthaltenden Elemente definiert, sondern durch den Grad ihrer Zugehörigkeit zu dieser Menge[33].
- **Künstliche Neuronale Netze (KNN)**: Aufbau analog zu biologischen neuronalen Netzen. Informationen werden von „Neuronen" verarbeitet und über gewichtete Verbindungen weitergegeben[34].
- **Relationale Operatoren**: Kombination verschiedener Datenquellen, die in Form von Tabellen vorliegen, auf Basis von Schlüsselattributen[35].

[29] s. Bleiholder u. Naumann 2011, S. 61

[30] s. Beyerer et al. 2006, S. 25; s. Ruser u. Puente León 2007, S. 98

[31] s. Beyerer et al. 2006, S. 25

[32] s. Ruser u. Puente León 2007, S. 99; s. Dietmayer 2006, S. 39 ff.

[33] s. Ruser u. Puente León 2006, S. 11; 2007, S. 99; s. Lehmann et al. 1992, S. 1 f.

[34] s. Fritsch u. Finke 2012, S. 307; vgl. Lawrence et al. 2012

[35] s.Bleiholder u. Naumann 2008, S. 5

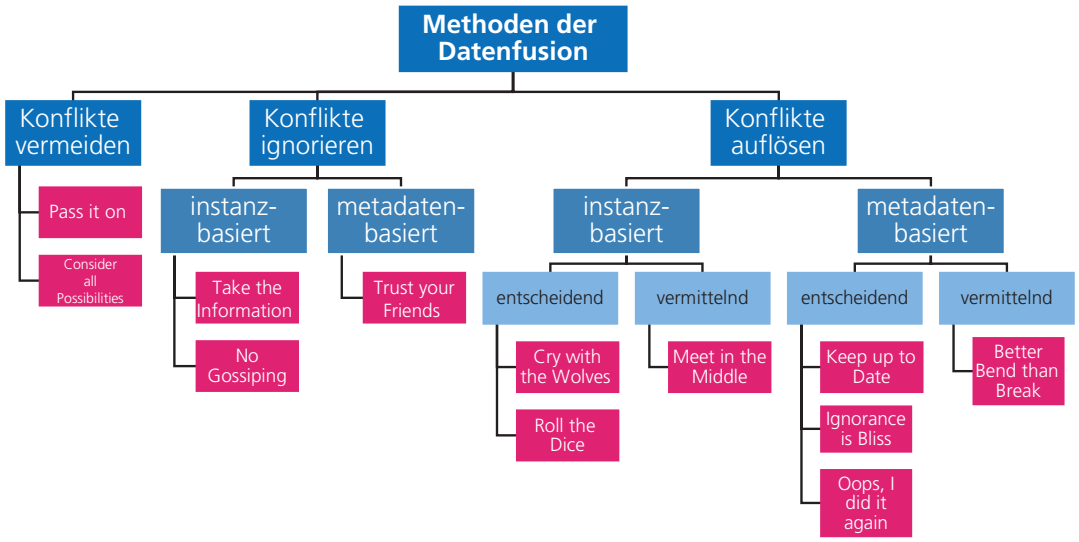


*Bild 12: Klassifizierung der Methoden der Datenfusion II (Bleiholder u. Naumann 2006, S. 4)*

Es lassen sich konkrete Entscheidungsregeln formulieren, die nach der Klassifikation zur Einteilung von Datenfusionsmethoden (s. Bild 2, S. 9) eingeordnet werden können. Die Einteilung einer Auswahl an relevanten Entscheidungsregeln ist in Bild 12 dargestellt. Auf diese soll hier anhand zweier Beispiele genauer eingegangen werden. Eine vollständige Beschreibung der einzelnen Entscheidungsregeln kann dem Glossar in Kapitel 8.5.1 (s. S. 51) entnommen werden.

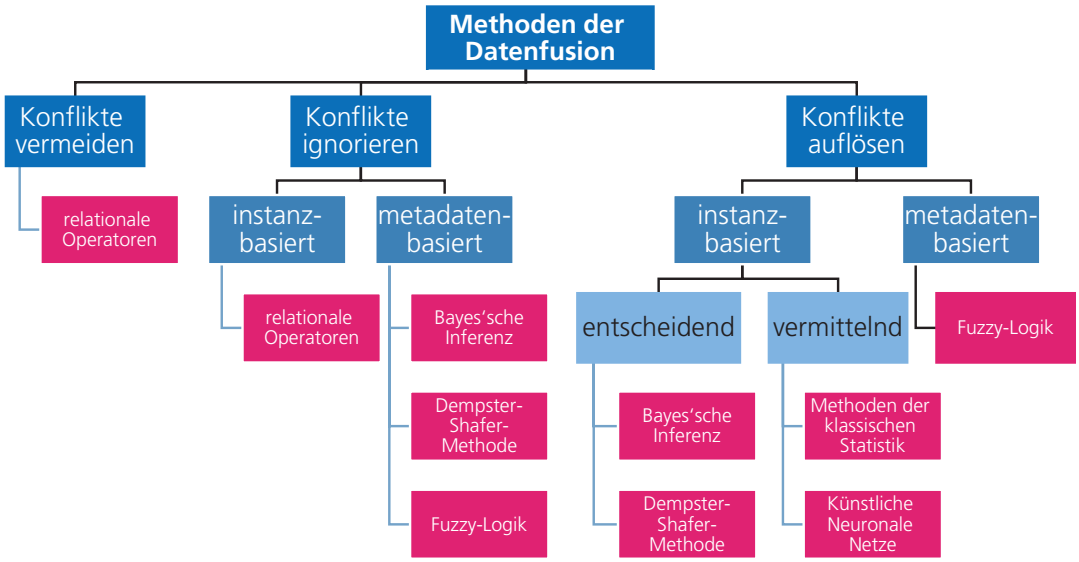


*Bild 13: Klassifizierung der Methoden der Datenfusion II (eigene Darstellung)*

Die Methoden der Datenfusion lassen sich wie bereits erwähnt in die Kategorien "Konflikte vermeiden", "Konflikte ignorieren" und "Konflikte auflösen" unterteilen. Die Methoden "Pass it on" und "Consider all Possibilities" gehören dabei zu den konfliktvermeidenden Methoden. Nach der Entscheidungsregel „Pass it on" werden konfligierende Attributwerte gemeinsam übernommen. Es wird dem Anwendenden des fusionierten Datenbestands überlassen, wie die auftretenden Datenkonflikte behandelt werden. Nach der Entscheidungsregel „Consider all Possibilities" werden alle möglichen Kombinationen der konfligierenden Attributwerte erzeugt und in den fusionierten Datenbestand übernommen. Problematisch ist hier die Erzeugung von Attributwerten, die nicht in den Ausgangsbeständen enthalten sind.

Ebenfalls ist es möglich, alle weiteren der oben genannten Datenfusionsmethoden in dieses Klassifikationsmodell einzuordnen. Die entsprechende Zuordnung kann Bild 13 entnommen werden.

Die Methoden der klassischen Statistik gehören zu den vermittelnden, instanzbasierten Konfliktlösungsstrategien. Die Bayes'sche Inferenz ist sowohl als metadatenbasierte, Konflikte ignorierende als auch als entscheidende, Konflikte auflösende Strategie anwendbar. Weiterhin ist die Dempster-Shafer-Methode sowohl als metadatenbasierte, Konflikte ignorierende als auch als entscheidende, Konflikte auflösende Strategie implementierbar. Die Fuzzy-Logik ist als metadatenbasierte, Konflikte ignorierende sowie als Konflikte auflösende Strategie anwendbar. Die Anwendung Künstlicher Neuronaler Netze wird den vermittelnden, Konflikte auflösenden Methoden der Datenfusion zugeordnet. Schließlich werden relationale Operatoren den konfliktvermeidenden oder instanzbasierten, Konflikte ignorierenden Methoden der Datenfusion zugeordnet. Wohlgemerkt handelt es sich bei den relationalen Operatoren um die einzigen Methoden der Datenfusion, welche nicht zur Auflösung von Konflikten angewendet werden können.

Informationen hinsichtlich der Anwendung dieser Methoden können dem Kapitel 8.5 im Glossar sowie dem Nachschlagewerk auf der Website dieses Projekts entnommen werden. Dort werden außerdem die im Folgenden in Bild14 (s. S. 23) fokussierten Vor- und Nachteile vollständig und detailliert erläutert.

Bei der Bewertung der einzelnen Datenfusionsmethoden lässt sich für die Entscheidungsregeln beispielsweise als Vorteil beschreiben, dass sie aufgrund ihres generischen Charakters für viele Anwendungsfälle konkretisierbar sind. Sie weisen zudem in der Regel einen geringen Implementierungsaufwand und liefern intuitiv nachvollziehbare Ergebnisse. Im Gegensatz dazu ist die Anwendung von Entscheidungsregeln nicht oder nur zum Teil automatisierbar. Die Vielzahl möglicher Strategien (s. Bild 12, S. 21) erschwert außerdem in einigen Fällen die Auswahl der am besten geeigneten. Weiterhin eignen sich die teilweise eher pauschal formulierten Strategien weniger für komplexe und stark differenzierte Problemstellungen. Die Vor- und Nachteile aller weiteren Datenfusionsmethoden können der nachfolgenden Tabelle entnommen werden.

[36] dafuer-tool.fir.de

Auf Basis des Bildes 14 können die optimalen Anwendungsgebiete für die jeweiligen Methoden formuliert werden. So eignen sich **Entscheidungsregeln** für die Anwendung weniger komplexer Systeme mit einer geringen Anzahl von zu fusionierenden Datenquellen[37]. Die Methoden der **klassischen Statistik** finden dort verstärkt Anwendung, wo umfangreiche Datensätze mit stochastisch verteilten Daten fusioniert werden sollen[38]. Die **Bayes'sche Inferenz** eignet sich besonders für Anwendungsfälle, bei denen Vorwissen über die Zuverlässigkeit der betrachteten Datenquellen verfügbar ist[39]. Mit der möglichen Modellierung von Unsicherheiten als Alleinstellungsmerkmal eignet sich die **Dempster-Shafer-Methode** insbesondere für die Fusion unzuverlässiger Datenquellen[40]. Ein geeignetes Anwendungsgebiet für die **Fuzzy-Logik** ist die Fusion von Daten, die Repräsentationen einer linguistischen Variablen darstellen und/oder mit Ungewissheit behaftet sind, sowie Problemstellungen, bei denen menschliches Vorwissen zu explizieren ist[41]. **Künstliche Neuronale Netze** bieten sich besonders für die Fusion von Klassifikationsergebnissen und sehr komplexen Problemen an, die mit den sonstigen Methoden der Datenfusion nur schwer zu modellieren sind[42]. Schließlich werden **relationale Operatoren** verstärkt für die Fusion von sich gegenseitig subsumierenden und/oder komplementierenden Datensätzen angewendet. Dies ist insbesondere der Fall, wenn verschiedene Aspekte desselben Objekts erfasst und aggregiert werden sollen.[43]

Im Anschluss an diesen Arbeitsschritt wird jetzt auf Basis von Bild 14 die Resistenz der verschiedenen Methoden gegenüber den abgeleiteten prozesstypischen Fehlern exemplarisch für eine Auswahl von Fehlerarten bewertet. Für die Herausforderungen eines unterschiedlichen Datenschemas, einer unterschiedlichen Syntax bei gleicher Semantik sowie einer unterschiedlichen Semantik bei gleicher Syntax der Daten ist die Eignung der verschiedenen Methoden der Datenfusion in Bild 15 (s. S. 24) dargestellt. Ein vollständig ausgefüllter Kreis beispielweise in der Zeile "Unterschiedliches Datenschema" impliziert dabei eine sehr hohe Resistenz der jeweiligen Datenfusionsmethode gegenüber der

[37] s. Bleiholder u. Naumann 2011, S. 61
[38] vgl. Jirak et al. 2018; s. Myung 2003, S. 91
[39] s. Beyerer et al. 2006, S. 24 f.; s. Ruser u. Puente León 2007, S. 99
[40] s. Durrant-Whyte u. Henderson 2008, S. 589 f.
[41] vgl. Rommelfanger 1993, S. 32
[42] s. Nelles 2006, S. 93 f.
[43] s. Bleiholder u. Naumann 2008, S. 20 f.

| | Vorteile | Nachteile |
|---|---|---|
| **Entscheidungsregeln** | • für viele Anwendungsfälle konkretisierbar<br>• geringer Implementierungsaufwand<br>• intuitiv nachvollziehbares Ergebnis | • nur teilweise automatisierbar<br>• nicht für komplexe Probleme geeignet<br>• durch Fülle an Strategien kein klar definiertes Vorgehen |
| **klassische Statistik** | • fundierte theoretische Grundlagen<br>• Erkenntnisse über die zugrundeliegende Wahrscheinlichkeitsverteilung<br>• universelle Anwendbarkeit | • A-priori-Wissen über Art der Wahrscheinlichkeitsverteilung notwendig<br>• anfällig gegenüber statistischen Ausreißern<br>• hohe Komplexität für multivariate Probleme |
| **Bayes'sche Inferenz** | • Einbeziehung von A-priori-Wissen<br>• explizite Berücksichtigung der Datenquelle | • A-priori-Wissen über Hypothesen notwendig<br>• Notwendigkeit sich gegenseitig ausschließender Hypothesen |
| **Dempster-Shafer-Methode** | • Modellierung von Unsicherheit<br>• Berücksichtigung unbekannter Ursachen<br>• Berücksichtigung von Plausibilität | • hoher Rechenaufwand<br>• differenzierte Darstellung des vorhandenen Wissens notwendig<br>• Undefiniertheit bei völlig gegensätzlichen Aussagen |
| **Fuzzy-Logik** | • Modellierung von Ungewissheit<br>• einfache und realitätsnahe Modellierung<br>• Modellierung von menschlichem Wissen | • keine Abbildung von Wahrscheinlichkeiten<br>• Zugehörigkeitsfunktion schwer validierbar<br>• hohe Subjektivität der Zugehörigkeitsfunktion |
| **Künstliche Neuronale Netze** | • verschiedenste Daten fusionierbar<br>• komplexe und nichtlineare Systeme modellierbar<br>• Zugriff auf vortrainierte Modelle | • keine Erkenntnisse über zugrundeliegende Zusammenhänge<br>• Fachkenntnisse erforderlich<br>• Trainingsdaten erforderlich |
| **relationale Operatoren** | • gut mit anderen Strategien kombinierbar<br>• einfache Implementierung<br>• als SQL-Anfrage implementierbar | • keine Datenkonflikte behandelbar<br>• Trade-off zwischen Vollständigkeit und Prägnanz |

*Bild 14: Vor- und Nachteile der verschiedenen Methoden der Datenfusion (eigene Darstellung)*

Fusion von Datensätzen mit unterschiedlichem Datenschema. Umgekehrt impliziert ein vollständig leerer Kreis in der gleichen Zeile eine geringe Resistenz derselben Datenfusionsmethode gegenüber dem genannten Fehler.

So eignen sich beispielsweise für die Fusion von Datensätzen mit unterschiedlichen Schemata Entscheidungsregeln nur bedingt, da viele Strategien ein zuverlässiges Schema-Matching voraussetzen. Die Methoden der klassischen Statistik, die Bayes'sche Inferenz und die Dempster-Shafer-Methode setzen ein solches Schema-Matching nicht zwangsläufig voraus. Mit der Fuzzy-Logik lassen sich die einzelnen Elemente eines Datentupels unabhängig von der übergeordneten Struktur bewerten. Künstliche Neuronale Netze sind sehr gut zur Fusion von Datensätzen mit unterschiedlichen Schemata geeignet, solange sich die jeweiligen Schemata nicht über die Zeit ändern und somit den Schemata der Trainingsdatensätze entsprechen. Relationale Operatoren zuletzt sind ebenfalls sehr gut geeignet, da mit einer Teilgruppe der Operatoren sämtliche Attributwerte aus beiden Quellen übernommen werden können.

Eine Gesamtübersicht über die Eignung der verschiedenen Methoden der Datenfusion für alle abgeleiteten Prozessfehler kann den Bildern 37 (s. S. 42) und 38 (s. S. 43) im Anhang entnommen werden.

Die Anwendenden sind mit dem vorgestellten Verfahren in der Lage, anhand der von ihnen ausgewählten Datenquellen die eigenen spezifischen Herausforderungen bei der Datenfusion zu identifizieren. Für die finale Auswahl einer für den Anwendungsfall geeigneten Datenfusionsmethode wurden für die ermittelten Prozessfehler die jeweiligen Eignungen der verschiedenen Methoden bewertet. Auf Grundlage dieser Bewertung kann nun diejenige Methode ausgewählt werden, die für die identifizierten Herausforderungen am besten geeignet ist.

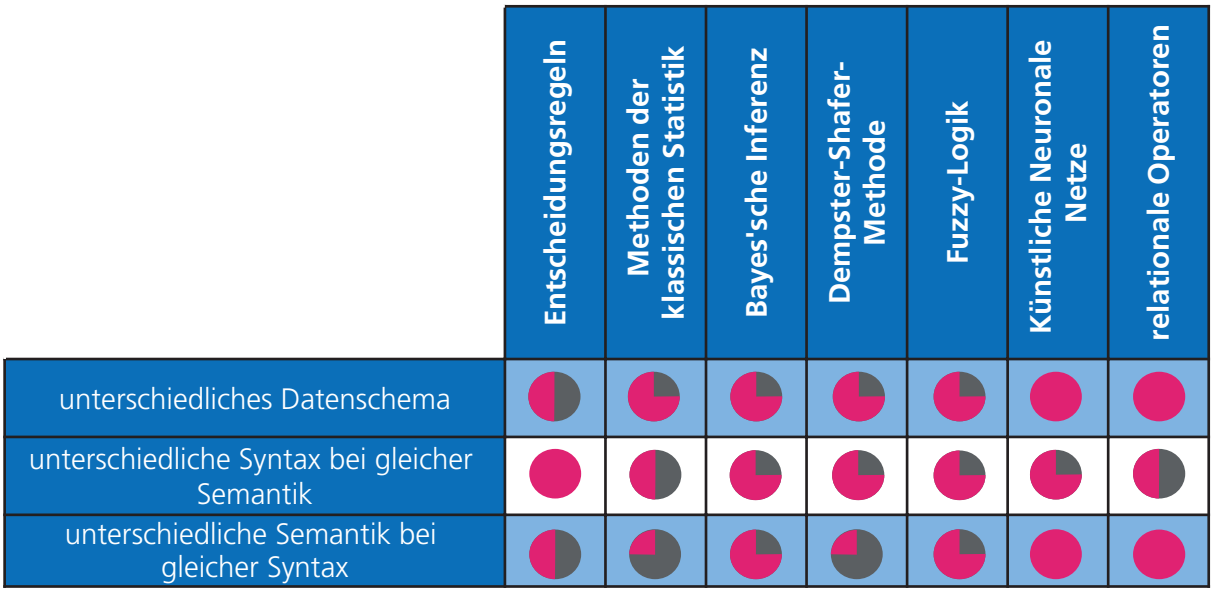


*Bild 15: Exemplarische Bewertung der Methoden der Datenfusion (Auszug) (eigene Darstellung)*

# 5 Zusammenfassung und Ausblick

Eine ausreichende Datenqualität und eine anwendungsgerechte Informationsverfügbarkeit sind notwendige Bedingungen für eine zuverlässige Entscheidungsfindung in der Produktionsplanung und -steuerung. Weiterhin wird für eine datenbasierte Wertschöpfung durch Verfahren wie das Data-Mining, eine vollständige und hochqualitative Datenbasis benötigt. Die Datenqualität lässt sich anhand verschiedener Qualitätsmerkmale definieren und quantifizieren: Eine Möglichkeit zur Steigerung der Datenqualität ist die Integration verschiedener Datenquellen. Die Datenfusion ist ein Prozessschritt der Datenintegration zur Behandlung von Konflikten innerhalb der zu integrierenden Daten. Zentrales Hemmnis bei der Anwendung der Datenfusion in Unternehmen ist die fehlende Möglichkeit der (teil-)automatischen Ableitung von geeigneten Fusionsmethoden. Dafür wurde im Zuge des Forschungsprojekts ‚DaFuER' ein Verfahren entwickelt, das in Abhängigkeit des gegebenen Anwendungsfalls die geeignetste der genannten Methoden der Datenfusion identifiziert. In einem ersten Schritt wird der Anwendungsfall definiert, dann werden die zu fusionierenden Datenquellen auf Basis der Zuordnung von Datenquellen zu Informationsbedarfen und ihren jeweiligen Datenqualitäten ausgewählt. Abschließend erfolgt die Auswahl geeigneter Methoden der Datenfusion in Abhängigkeit der vorliegenden prozesstypischen Fehler.

Die Idee dieses Leitfadens war es, die Forschungsergebnisse, insbesondere Prozesse, gesammelte Datenquellen und Methoden der Datenfusion, zusammenfassend zu beschreiben und eine Auswahlhilfe für Unternehmen sowie für Entwickler*innen betrieblicher Anwendungssysteme zu bieten. Hinsichtlich der Unternehmen lag ein besonderer Fokus auf der Schaffung von Nutzenvorteilen für kleine und mittlere Unternehmen, weshalb diese bereits frühzeitig in das Projekt integriert wurden. So wurde das allgemeine Vorgehen zur Anwendung der Datenfusion bereits während der Entwicklung bei mehreren Projektpartnern anhand von realen Anwendungsfällen erprobt und gewonnene Erkenntnisse zur Optimierung der eigenen Abläufe in den Produktionsplanungen erfolgreich integriert.

# 6 Das FIR als kompetenter Partner in der Praxis

Unser Ziel im Bereich Produktionsmanagement des FIR e. V. an der RWTH Aachen ist die Optimierung der Geschäftsprozesse, sodass Unternehmen die Effizienz ihrer Abläufe steigern oder neue Geschäftsbereiche entwickeln können, um ihre Wettbewerbsfähigkeit zu sichern und auszubauen. Im Zentrum der Betrachtung liegen dabei der industrielle Auftragsabwicklungsprozess und seine Teilprozesse. Dies umfasst den Vertrieb inklusive Angebotsklärung, Beschaffung sowie die Produktionsplanung und -steuerung bis hin zum Versand.

Neben den Informationsflüssen liegen die inner- und überbetrieblichen Materialflüsse im Fokus unserer Betrachtung. Dabei spielen die Kernelemente und Ziele der Digitalisierung, u. a. die digitale Prozessautomatisierung und -optimierung und die Sammlung und Auswertung von Daten, eine wesentliche Rolle, denn die steigende Verfügbarkeit von Technologien schafft neue Anwendungsszenarien. Neue Technologien beeinflussen die Architektur und Funktionalitäten etablierter betrieblicher Anwendungssysteme, etwa ERP-Systeme, und produktionsnaher Systeme wie Manufacturing-Execution-Systems. Wir untersuchen die Möglichkeiten der Verzahnung von logistischen Prozessen und Geschäftsprozessen sowie der dazugehörigen Informationssysteme innerhalb der Unternehmensorganisation und im Kunden-Lieferanten-Verhältnis. Für alle Geschäftsprozesse gleichermaßen erforschen wir außerdem neue Möglichkeiten im Umgang mit Stamm- und Bewegungsdaten, also der Steigerung der Datenqualität oder der Gewinnung neuer Erkenntnisse durch Verfahren der Datenanalyse. Für Fragen und Anmerkungen steht Ihnen unser Ansprechpartner gerne zur Verfügung:

## Kontakt

Jokim Janßen, M.Sc.
Projektmanager
FIR e. V. an der RWTH Aachen
Bereich Produktionsmanagement
Fachgruppe Supply-Chain-Management
Tel.: +49 241 47705-413
E-Mail: Jokim.Janssen@fir.rwth-aachen.de

# 7 Anhang

| | | | Losgrößenrechnung | Feinterminierung | Ressourcenfeinplanung | Reihenfolgeplanung | Verfügbarkeitsprüfung | Auftragsfreigabe |
|---|---|---|---|---|---|---|---|---|
| Stammdaten | Materialstammdaten | Materialnummer | x | x | x | x | x | x |
| | | Lagerinformationen | x | x | | | x | |
| | | Kosteninformationen | x | | | x | | |
| | Produktionsdaten | Losgröße | x | | | | | |
| | | Standard-Plan-Durchlaufzeit | x | x | x | | | |
| | | Standard-Arbeitsplan-Nummer | x | x | x | x | x | x |
| | | Arbeitsgangnummer | x | | | x | x | x |
| | | Gesamtbedarfsmenge | | | | x | x | x |
| | Arbeitsplandaten | Arbeitsplannummer | x | x | x | x | x | x |
| | | Arbeitsplan-Variantennummer | x | x | x | x | x | x |
| | | Arbeitsplatz | | | x | x | x | x |
| | | Zeitdauer Durchführung | x | x | x | x | x | x |
| | | benötigte Fertigungshilfsmittel | x | x | x | x | x | x |
| | | Belegungszeitfaktor | | x | x | | | |
| | Ressourcendaten | Maschinenkapazität | x | | x | x | | |
| | | Instandhaltungsdaten | | | x | | x | |
| | | Lagerbestände | x | | | x | | |
| | | Bedarfe | x | x | | | | x |
| Bewegungsdatem | Auftragsdaten | Auftragsnummer | | x | x | x | x | x |
| | | Auftragszeiten | | x | x | x | x | x |
| | | Arbeitsgangzeiten | | x | | | | x |
| | | Auftragsfortschritt | | x | | | | |
| | | Bearbeitungszeit | x | x | | x | x | |
| | | Transportzeit | | x | | x | | |
| | | Liegezeit | | x | | x | | |
| | | Rüstzeit | | x | | x | | |
| | | Durchlaufzeit | | x | | x | | |
| | | produzierte Menge | | x | x | x | | x |
| | | Ausschuss | | x | x | x | | |
| | | Störungen | | x | x | x | | |
| | | Maschinenzeiten | x | | x | | | |

*Bild 16: Informationsbedarfe der Produktionsplanung und -steuerung (eigene Darstellung)*

| | | Informationsbedarf | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Materialnummer | Lagerinformationen | Kosteninformationen | Losgröße | Standard-Plan-Durchlaufzeit | Standard-Arbeitsplan-Nummer | Arbeitsgangnummer | Gesamtbedarfsmenge | Arbeitsplannummer | Arbeitsplan-Variantennummer | Arbeitsplatz | Zeitdauer Durchführung | benötigte Fertigungshilfsmittel | Belegungszeitfaktor | Maschinenkapazität | Instandhaltungsdaten |
| Informationsverfügbarkeit | MDE | | | | | | | x | | | | x | x | x | | x | x |
| | Informationssysteme | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x |
| | Intelligente Sensorik | | | | | | | | | | | | x | | | | |
| | RTLS | | x | | | | | | x | | | x | | | x | | |
| | Auto-ID | x | x | x | | | | | x | | | | | | | | |
| | Lesegeräte | x | | | | | | | | | | x | | | | | |
| | Buzzer | | | | | | | | | | | | | | | | x |
| | PZE-Terminal | | | | | | | | | | | x | x | x | x | | |
| | HMI | x | | | | | | x | | | | x | x | x | x | x | x |
| | Waagen | | x | | | | | | | | | | | | | | |
| | Begleitpapiere | x | x | x | x | x | x | x | x | x | x | | | | | | |
| | mechanische Zählwerke | | | | | | | | | | | | | | | | |
| | Mobile-Applications | x | x | x | | x | x | | x | x | x | | x | x | x | x | |
| | Internetquellen | x | x | x | | | | | | | | | | x | | x | |
| | BDE-Terminal | x | x | x | | | | | | x | | x | x | | x | | x |
| | Dokumentation | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x |
| | Fabrikkalender | | | | | x | x | | | x | x | | x | | | | |
| | Offline-Datenbank | x | x | x | x | x | x | | x | | x | x | | x | x | x | |
| | Schriftverkehr | x | x | x | x | x | x | | x | | x | x | | x | | | |
| | menschliches Wissen | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x |

*Bild 17: Zuordnung von Datenquellen zu Informationsbedarfen (Bewegungsdaten) (eigene Darstellung)*

| | Vollständigkeit | Aktualität | Fehlerfreiheit | Zugänglichkeit | Objektivität | Genauigkeit |
|---|---|---|---|---|---|---|
| MDE |  |  |  |  |  |  |
| Informationssysteme |  |  |  |  |  |  |
| Intelligente Sensorik |  |  |  |  |  |  |
| RTLS |  |  |  |  |  |  |
| Auto-ID |  |  |  |  |  |  |
| Lesegeräte |  |  |  |  |  |  |
| Buzzer |  |  |  |  |  |  |
| PZE-Terminal |  |  |  |  |  |  |
| HMI |  |  |  |  |  |  |
| Waagen |  |  |  |  |  |  |
| Begleitpapiere |  |  |  |  |  |  |
| mechanische Zählwerke |  |  |  |  |  |  |
| Mobile-Applications |  |  |  |  |  |  |
| Internetquellen |  |  |  |  |  |  |
| BDE-Terminal |  |  |  |  |  |  |
| Dokumentation |  |  |  |  |  |  |
| Fabrikkalender |  |  |  |  |  |  |
| Offline-Datenbank |  |  |  |  |  |  |
| Schriftverkehr |  |  |  |  |  |  |
| menschliches Wissen |  |  |  |  |  |  |

*Bild 18: Bewertung der Datenqualität der Datenquellen (eigene Darstellung)*

| | | | | |
|---|---|---|---|---|
| **Herkunft** | intern | | extern | |
| **Metadaten** | vorhanden | teilweise vorhanden | nicht vorhanden | |
| **Art der Erfassung** | automatisch | halbautomatisch | manuell | |
| **Erfassungs-auslösung** | kontinuierlich | zyklisch | getriggert | |
| **Archivierungs-zeitraum** | > 1 Monat | > 1 Woche | > 1 Tag | < 1 Tag |
| **Aktualisierungs-rate** | > 1 Stunde | < 1 Stunde | < 1 Minute | < 1 Sekunde |
| **Schnittstellen** | Maschinen-schnittstelle | Software-schnittstelle | Mensch-Maschine | Mensch-Mensch |
| **Datentyp** | Integer | Float | Boolean | String |
| **Datenstruktur** | strukturiert | semistrukturiert | unstrukturiert | |
| **Skalenniveau** | Nominal-skala | Ordinal-skala | Intervall-skala | Verhältnis-skala |
| **Auflösung** | > 10 % | < 10 % | < 1 % | < 0,1 % |
| **Störanfälligkeit** | hoch | mittel | gering | |

*Bild 19: Morphologie – Informationssysteme (eigene Darstellung)*

| | | | | |
|---|---|---|---|---|
| **Herkunft** | intern | | extern | |
| **Metadaten** | vorhanden | teilweise vorhanden | nicht vorhanden | |
| **Art der Erfassung** | automatisch | halbautomatisch | manuell | |
| **Erfassungs-auslösung** | kontinuierlich | zyklisch | getriggert | |
| **Archivierungs-zeitraum** | > 1 Monat | > 1 Woche | > 1 Tag | < 1 Tag |
| **Aktualisierungs-rate** | > 1 Stunde | < 1 Stunde | < 1 Minute | < 1 Sekunde |
| **Schnittstellen** | Maschinen-schnittstelle | Software-schnittstelle | Mensch-Maschine | Mensch-Mensch |
| **Datentyp** | Integer | Float | Boolean | String |
| **Datenstruktur** | strukturiert | semistrukturiert | unstrukturiert | |
| **Skalenniveau** | Nominal-skala | Ordinal-skala | Intervall-skala | Verhältnis-skala |
| **Auflösung** | > 10 % | < 10 % | < 1 % | < 0,1 % |
| **Störanfälligkeit** | hoch | mittel | gering | |

*Bild 20: Morphologie – Intelligente Sensorik (eigene Darstellung)*

| | | | | |
|---|---|---|---|---|
| **Herkunft** | intern | | extern | |
| **Metadaten** | vorhanden | teilweise vorhanden | nicht vorhanden | |
| **Art der Erfassung** | automatisch | halbautomatisch | manuell | |
| **Erfassungs-auslösung** | kontinuierlich | zyklisch | getriggert | |
| **Archivierungs-zeitraum** | > 1 Monat | > 1 Woche | > 1 Tag | < 1 Tag |
| **Aktualisierungs-rate** | > 1 Stunde | < 1 Stunde | < 1 Minute | < 1 Sekunde |
| **Schnittstellen** | Maschinen-schnittstelle | Software-schnittstelle | Mensch-Maschine | Mensch-Mensch |
| **Datentyp** | Integer | Float | Boolean | String |
| **Datenstruktur** | strukturiert | semistrukturiert | unstrukturiert | |
| **Skalenniveau** | Nominal-skala | Ordinal-skala | Intervall-skala | Verhältnis-skala |
| **Auflösung** | > 10 % | < 10 % | < 1 % | < 0,1 % |
| **Störanfälligkeit** | hoch | mittel | gering | |

*Bild 21: Morphologie – RTLS (eigene Darstellung)*

Eine vollständige Abbildung der Morphologie aller im Rahmen dieses Leitfadens betrachteten Datenquellen kann der Website des Projekts unter dafuer-tool.fir.de entnommen werden.

| | MDE | Informations-systeme | Intelligente Sensorik | RTLS | Auto-ID |
|---|---|---|---|---|---|
| Herkunft | intern | | | | |
| Metadaten | vorhanden | vorhanden | teilweise vorhanden | vorhanden | vorhanden |
| Art der Erfassung | automatisch | | | | |
| Erfassungs-auflösung | kontinu-ierlich, zyklisch, getriggert | zyklisch, getriggert | kontinu-ierlich, getriggert | zyklisch | getriggert |
| Archivierungs-zeitraum | > 1 Monat | > 1 Monat | > 1 Woche | < 1 Tag | < 1 Tag |
| Aktualisie-rungsrate | < 1 Sekunde | < 1 Minute | < 1 Sekunde | < 1 Minute | < 1 Minute |
| Schnittstellen | Software-, Maschinen-schnittstelle | Software-schnittstelle | Software-, Maschinen-schnittstelle | Software-schnittstelle | Software-schnittstelle |
| Datentyp | Float, Boolean | Integer, Float, Boolean, String | Float, Boolean | Float | Integer, String |
| Datenstruktur | strukturiert | | | | |
| Skalenniveau | Verhältnis-skala | Nominal-Intervall-, Verhältnis-skala | Verhältnis-skala | Intervall-skala | Nominal-, Intervall-skala |
| Auflösung | < 0,1% | < 1% | < 1%, < 0,1% | < 1%, < 0,1% | < 1%, < 0,1% |
| Störanfällig-keit | gering | mittel | mittel | mittel | mittel |

*Bild 22: Morphologische Ausprägungen von Datenquellen – Datenquellen mit automatischer Datenerfassung (eigene Darstellung)*

| | Lesegeräte | Buzzer | PZE-Terminal | HMI | Waagen |
|---|---|---|---|---|---|
| Herkunft | intern | | | | |
| Metadaten | vorhanden | nicht vorhanden | vorhanden | vorhanden | vorhanden |
| Art der Erfassung | halbautomatisch | | | | |
| Erfassungs-auflösung | getriggert | getriggert | getriggert | kontinu-ierlich | getriggert |
| Archivierungs-zeitraum | > 1 Tag | < 1 Tag | > 1 Monat | < 1 Tag | < 1 Tag |
| Aktualisie-rungsrate | < 1 Stunde | > 1 Stunde | > 1 Stunde | < 1 Sekunde | < 1 Minute |
| Schnittstellen | Software-schnitt-stelle, Mensch-Maschine | Maschinen-schnittstelle | Software-, Maschinen-schnittstelle | Mensch-Maschine | Maschinen-schnittstelle Mensch-Maschine |
| Datentyp | Integer, String | Boolean | Integer | Float, Boolean, String | Float |
| Datenstruktur | strukturiert | | | | |
| Skalenniveau | Nominal-, Intervall-skala | Nominal-skala | Intervall-skala | Verhältnis-skala | Verhältnis-skala |
| Auflösung | < 10%, < 1% | > 10% | < 0,1 % | < 0,1 % | < 0,1% |
| Störanfälligkeit | mittel | mittel | mittel | gering | gering |

*Bild 23: Morphologische Ausprägungen von Datenquellen – Datenquellen mit halbautomatischer Datenerfassung (1/2) (eigene Darstellung)*

| | Begleit-papiere | Mechanische Zählwerke | Mobile-Applications | Internet-quellen |
|---|---|---|---|---|
| Herkunft | intern, extern | intern | intern, extern | extern |
| Metadaten | teilweise vorhanden | nicht vorhanden | vorhanden | teilweise vorhanden |
| Art der Erfassung | halbauto-matisch | halbauto-matisch | halbauto-matisch | halbauto-matisch |
| Erfassungs-auflösung | zyklisch | getriggert | getriggert | getriggert |
| Archivierungs-zeitraum | > 1 Tag | < 1 Tag | > 1 Monat | > 1 Monat |
| Aktualisierungs-rate | > 1 Stunde | < 1 Sekunde | < 1 Sekunde | < 1 Sekunde |
| Schnittstellen | Mensch - Maschine | Mensch - Maschine | Software - schnittstelle , Mensch - Maschine | Software - schnittstelle |
| Datentyp | Float, String | Integer | Float , Boolean, String | Float , String |
| Datenstruktur | strukturiert | strukturiert | strukturiert | strukturiert, unstrukturiert |
| Skalenniveau | Intervallskala | Verhältnis-skala | Intervallskala | Nominal -, Ordinal -, Intervall -, Verhältnis - skala |
| Auflösung | < 1% | < 1% | < 1% | < 10%, < 1% |
| Störanfälligkeit | mittel | gering | mittel | hoch |

*Bild 24: Morphologische Ausprägungen von Datenquellen – Datenquellen mit halbautomatischer Datenerfassung (2/2) (eigene Darstellung)*

| | BDE-Terinal | Dokumen-tation | Fabrik-kalender | Offline-Datenbank | Schrift-verkehr | mensch-liches Wissen |
|---|---|---|---|---|---|---|
| Herkunft | intern | intern, extern | intern | intern | intern, extern | intern, extern |
| Metadaten | vorhan-den | teilweise vorhan-den | vorhan-den | teilweise vorhan-den | nicht vorhan-den | nicht vorhan-den |
| Art der Erfassung | manuell | | | | | |
| Erfassungs-auflösung | getriggert | getriggert | zyklisch | getriggert | getriggert | getriggert |
| Archivie-rungs-zeitraum | > 1 Monat | > 1 Monat | > 1 Monat | > 1 Woche | > 1 Monat, > 1 Woche | > 1 Monat |
| Aktualisie-rungsrate | < 1 Stunde | > 1 Stunde | > 1 Stunde | > 1 Stunde | > 1 Stunde | > 1 Stunde |
| Schnitt-stellen | Software-schnitt-stelle, Mensch-Maschine | Mensch-Maschine | Software-schnitt-stelle, Mensch-Maschine | Software-schnitt-stelle | Mensch-Mensch | Mensch-Mensch |
| Datentyp | Float, Boolean | Integer, Boolean, String | Integer, String | Float, String | String | String |
| Daten-struktur | struktu-riert | struktu-riert | struktu-riert | struktu-riert, semi-struktu-riert | unstruktu-riert | unstruktu-riert |
| Skalen-niveau | Intervall-skala | Intervall-skala | Intervall-skala | Intervall-skala, Verhält-nisskala | Nominal-skala | Nominal-, Ordinal-, Intervall-, Verhält-nisskala |
| Auflösung | < 1% | < 10% | < 1% | > 10%, < 1% | > 10% | > 10%, < 10%, < 1% |
| Störanfällig-keit | mittel | mittel | gering | mittel | hoch | hoch |

*Bild 25: Morphologische Ausprägung von Datenquellen – Datenquellen mit manueller Datenerfassung (eigene Darstellung)*

| | vorhanden | teilweise vorhanden | nicht vorhanden |
|---|---|---|---|
| vorhanden | | | |
| teilweise vorhanden | | | Metadaten nur für eine Quelle verfügbar |
| nicht vorhanden | | Metadaten nur für eine Quelle verfügbar | keine Metadaten verfügbar |

*Bild 26: Ableitung prozesstypischer Fehler – Metadaten (eigene Darstellung)*

| | automatisch | halbautomatisch | manuell |
|---|---|---|---|
| automatisch | potenziell große Datenmengen | | unterschiedliche Syntax |
| halbautomatisch | | | Erfassung unterschiedlicher Objekte |
| manuell | unterschiedliche Syntax | Erfassung unterschiedlicher Objekte | unterschiedliche Zuverlässigkeit |

*Bild 27: Ableitung prozesstypischer Fehler – Art der Erfassung (eigene Darstellung)*

| | > 1 Monat | > 1 Woche | > 1 Tag | < 1 Tag |
|---|---|---|---|---|
| > 1 Monat | | | historische Daten eingeschränkt verfügbar | historische Daten eingeschränkt verfügbar |
| > 1 Woche | | | historische Daten eingeschränkt verfügbar | historische Daten eingeschränkt verfügbar |
| > 1 Tag | historische Daten eingeschränkt verfügbar | historische Daten eingeschränkt verfügbar | keine historischen Daten verfügbar | keine historischen Daten verfügbar |
| < 1 Tag | historische Daten eingeschränkt verfügbar | historische Daten eingeschränkt verfügbar | keine historischen Daten verfügbar | keine historischen Daten verfügbar |

*Bild 28: Ableitung prozesstypischer Fehler – Archivierungszeitraum (eigene Darstellung)*

| | kontinuierlich | zyklisch | getriggert |
|---|---|---|---|
| kontinu-ierlich | potenziell große Datenmengen | potenziell große Datenmengen | unterschiedlich viele Datensätze pro Objekt |
| zyklisch | potenziell große Datenmengen | | unterschiedlich viele Datensätze pro Objekt |
| getriggert | unterschiedlich viele Datensätze pro Objekt | unterschiedlich viele Datensätze pro Objekt | Erfassung unterschiedlicher Objekte |

*Bild 29: Ableitung prozesstypischer Fehler – Erfassungsauflösung (eigene Darstellung)*

| | > 1 Stunde | < 1 Stunde | < 1 Minute | < 1 Sekunde |
|---|---|---|---|---|
| **> 1 Stunde** | keine aktuellen Daten verfügbar | | unterschiedliche zeitliche Auflösung | unterschiedliche zeitliche Auflösung |
| **< 1 Stunde** | | | unterschiedliche zeitliche Auflösung | unterschiedliche zeitliche Auflösung |
| **< 1 Minute** | unterschiedliche zeitliche Auflösung | unterschiedliche zeitliche Auflösung | | unterschiedliche zeitliche Auflösung |
| **< 1 Sekunde** | unterschiedliche zeitliche Auflösung | unterschiedliche zeitliche Auflösung | unterschiedliche zeitliche Auflösung | |

*Bild 30: Ableitung prozesstypischer Fehler – Aktualisierungsrate (eigene Darstellung)*

| | strukturiert | semistrukturiert | unstrukturiert |
|---|---|---|---|
| **strukturiert** | | | fehlende Schlüsselattribute, eingeschränktes Schema-Mapping |
| **semistruk-turiert** | | | fehlende Schlüsselattribute, eingeschränktes Schema-Mapping |
| **unstruk-turiert** | fehlende Schlüsselattribute, eingeschränktes Schema-Mapping | fehlende Schlüsselattribute, eingeschränktes Schema-Mapping | fehlende Schlüsselattribute, eingeschränktes Schema-Mapping |

*Bild 31: Ableitung prozesstypischer Fehler – Datenstruktur (eigene Darstellung)*

| | Maschinen-schnitstelle | Software-schnittstelle | Mensch-Maschine | Mensch-Mensch |
|---|---|---|---|---|
| Maschinen-schnittstelle | potenziell große Datenmengen | potenziell große Datenmengen | | unterschiedliche Syntax |
| Software-schnittstelle | potenziell große Datenmengen | potenziell große Datenmengen | unterschiedliche Schemata | unterschiedliche Syntax |
| Mensch-Maschine | | unterschiedliche Schemata | | |
| Mensch-Mensch | unterschiedliche Syntax | unterschiedliche Syntax | | hohe Subjektivität |

*Bild 32: Ableitung prozesstypischer Fehler – Schnitttstellen (eigene Darstellung)*

| | hoch | mittel | gering |
|---|---|---|---|
| hoch | hohe Varianz in Datensätzen | | unterschiedliche Zuverlässigkeit der Datenquellen |
| mittel | | | |
| gering | unterschiedliche Zuverlässigkeit der Datenquellen | | |

*Bild 33: Ableitung prozesstypischer Fehler – Störanfälligkeit (eigene Darstellung)*

| | Integer | Float | Boolean | String |
|---|---|---|---|---|
| Integer | | | unterschiedliche Syntax, unterschiedliche Semantik | keine mathematischen Operatoren anwendbar |
| Float | | | unterschiedliche Syntax, unterschiedliche Semantik | keine mathematischen Operatoren anwendbar |
| Boolean | unterschiedliche Syntax, unterschiedliche Semantik | unterschiedliche Syntax, unterschiedliche Semantik | | keine mathematischen Operatoren anwendbar |
| String | keine mathematischen Operatoren anwendbar | keine mathematischen Operatoren anwendbar | keine mathematischen Operatoren anwendbar | |

*Bild 34: Ableitung prozesstypischer Fehler – Schnitttstellen (eigene Darstellung)*

| | Nominalskala | Ordinalskala | Intervallskala | Verhältnisskala |
|---|---|---|---|---|
| Nominal-skala | keine mathematischen Operatoren anwendbar | keine mathematischen Operatoren anwendbar | unterschiedliches Skalenniveau | unterschiedliches Skalenniveau |
| Ordnial-skala | keine mathematischen Operatoren anwendbar | keine mathematischen Operatoren anwendbar | unterschiedliches Skalenniveau | unterschiedliches Skalenniveau |
| Intervall-skala | unterschiedliches Skalenniveau | unterschiedliches Skalenniveau | | unterschiedliches Skalenniveau |
| Verhältnis-skala | unterschiedliches Skalenniveau | unterschiedliches Skalenniveau | unterschiedliches Skalenniveau | |

*Bild 35: Ableitung prozesstypischer Fehler – Skalenniveau (eigene Darstellung)*

| | > 10 % | < 10 % | < 1% | < 0,1 % |
|---|---|---|---|---|
| > 10 % | geringe diskrete Auflösung | geringe diskrete Auflösung | überproportio-nale Gewichtung von Datenwerten | überproportio-nüle Gewichtung von Datenwerten |
| < 10 % | geringe diskrete Auflösung | geringe diskrete Auflösung | überproportio-nale Gewichtung von Datenwerten | überproportio-nale Gewichtung von Datenwerten |
| < 1 % | überproportio-nale Gewichtung von Datenwerten | überproportio-nale Gewichtung von Datenwerten | | |
| < 0,1 % | überproportio-nale Gewichtung von Datenwerten | überproportio-nale Gewichtung von Datenwerten | | |

*Bild 36: Ableitung prozesstypischer Fehler – Auflösung (eigene Darstellung)*

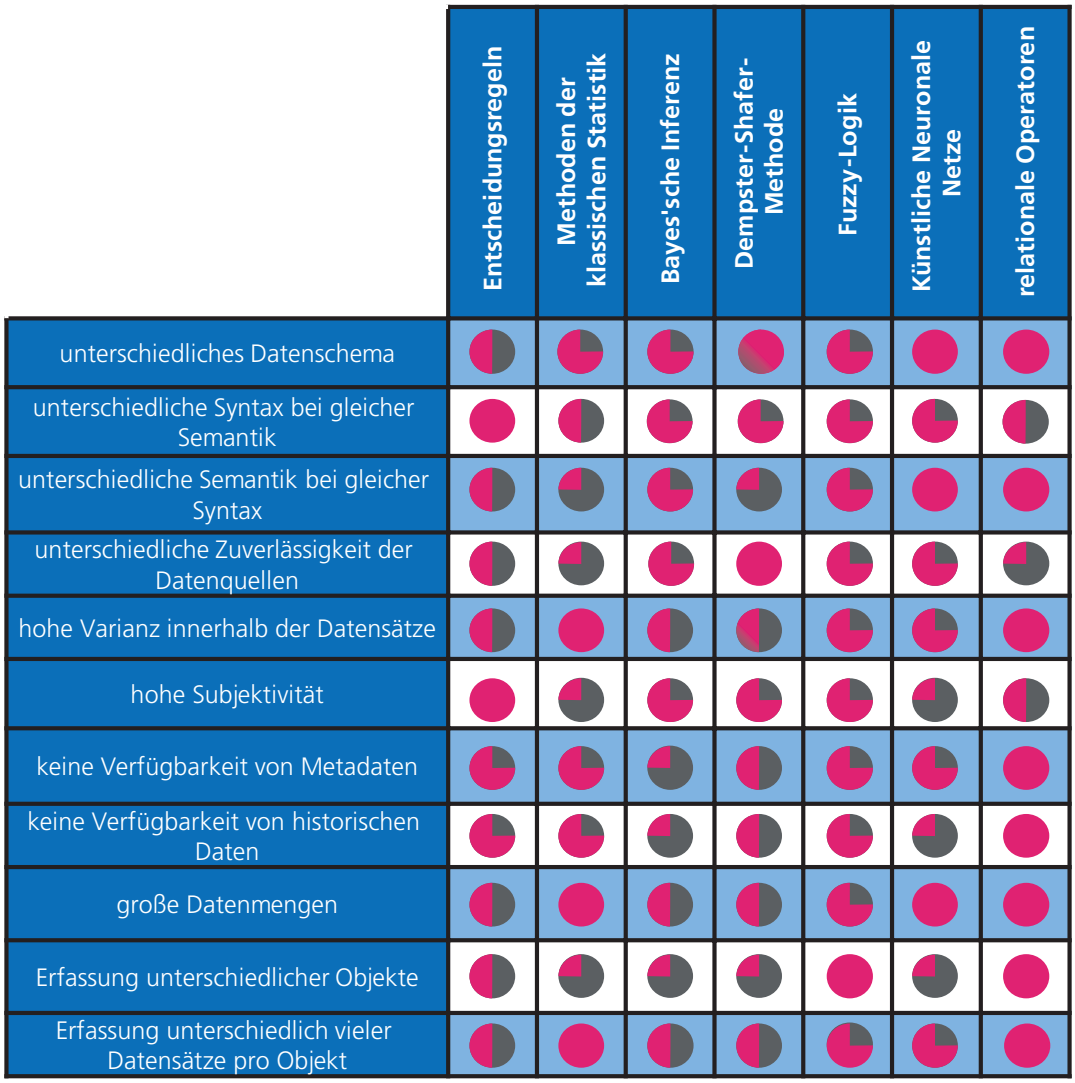


*Bild 37: Bewertung der Eignung von Methoden der Datenfusion (1/2) (eigene Darstellung)*


*Bild 38: Bewertung der Eignung von Methoden der Datenfusion (2/2) (eigene Darstellung)*

# 8 Glossar

Der folgende Abschnitt bietet eine Definitionssammlung der im Rahmen dieses Leitfadens relevanten Fachbegriffe, die vor allem das Verständnis für die Abbildungen der Kapitel 3 bis 5 erleichtern soll.

## 8.1 Informationsbedarfe der Produktionsplanung und -steuerung

Die hier aufgeführten Begriffe zur Charakterisierung des Informationsbedarfs der Produktionsplanung und -steuerung basieren auf den Definitionen des Glossars der SAP-Bibliothek, welche über folgenden Link aufgerufen werden kann:
help.sap.com/doc/saphelp_glossary/latest/de-DE/35/2cd77bd7705394e10000009b387c12/frameset.htm

| Klassifikation | Information | Beschreibung |
|---|---|---|
| Materialstammdaten | Materialnummer | Die Materialnummer ist eine Nummer, die einen Materialstammsatz und somit ein Material eindeutig identifiziert. |
| | Lagerinformationen | Lagerinformationen umfassen alle relevanten Informationen zur Lagerung eines Materials, wie z. B. den Raumbedarf pro Stück oder die maximale Lagerdauer. |
| | Kosteninformationen | Kosteninformationen umfassen alle Informationen, welche im Zuge der Verwendung eines Materials anfallen können, wie z. B. Eigenfertigungs-/ Bestellkosten, Lagerkosten oder Stückkostensätze. |
| Produktionsdaten | Losgröße | Die Losgröße ist die Menge einer Produktart oder Baugruppe, die in einer Produktionsstufe als geschlossener Posten (Los) ohne Unterbrechung gefertigt wird. |
| | Standard-Plan-Durchlaufzeit | Die Standard-Plan-Durchlaufzeit ist die Zeit, welche ein Arbeitsobjekt benötigt, um die standardmäßige Ablauffolge der Durchführung einer Aufgabe zu durchlaufen. |
| | Standard-Arbeitsplan-Nummer | Die Standard-Arbeitsplan-Nummer ist die Nummer eines Arbeitsplans, der eine Folge von Vorgängen festlegt, die sich häufig wiederholen. Standardarbeitspläne dienen als Vorlage für die Erstellung von Normalarbeitsplänen. Der Normalarbeitsplan wiederum ist ein Arbeitsplantyp, der eine oder mehrere Folgen von Vorgängen zur Fertigung eines Materials festlegt. Zur Verringerung des Erfassungsaufwands können in einen Normalarbeitsplan Standardarbeitspläne integriert werden. |
| | Arbeitsgangnummer | Die Arbeitsgangnummer ist die Nummer eines Arbeitsgangs, welche einen Schritt in einem Arbeitsplan angibt. |

| | | |
|---|---|---|
| | Gesamtbedarfsmenge | Die Gesamtbedarfsmenge ist die gesamte benötigte Menge eines bestimmten Produkts zur vollständigen Durchführung eines Produktionsprozesses. |
| Arbeitsplandaten | Arbeitsplannummer | Die Arbeitsplannummer ist die Nummer eines Arbeitsplans. |
| | Arbeitsplan-Variantennummer | Die Arbeitsplan-Variantennummer ist eine individuelle Nummer, welche einen Arbeitsplan als Variante eines übergeordneten Arbeitsplans kennzeichnet und ihn somit von anderen Varianten unterscheidet. |
| | Arbeitsplatz | Der Arbeitsplatz ist eine physische Einheit, die einen zweckmäßig eingerichteten räumlichen Bereich darstellt, in dem zugeordnete Vorgänge durchgeführt werden. |
| | Zeitdauer Durchführung | Die Zeitdauer der Durchführung ist die Zeitspanne zwischen Beginn des ersten Arbeitsgangs und Abschluss des letzten Arbeitsgangs eines Arbeitsplans. |
| | Benötigte Fertigungshilfsmittel | Benötigte Fertigungshilfsmittel umfassen benötigte nichtstationäre Betriebsmittel, welche in der Fertigung oder Instandhaltung eingesetzt werden. |
| Ressourcendaten | Maschinenkapazität | Die Maschinenkapazität beschreibt das Potenzial der Maschinen zum Ausstoß von Leistungen. |
| | Instandhaltungsdaten | Instandhaltungsdaten umfassen Daten, welche während Maßnahmen zum Erhalt der Funktionsfähigkeit technischer Systeme gesammelt werden. |
| | Lagerbestände | Lagerbestände umfassen die Menge eines oder mehrerer Materialien, welche aktuell physisch im Bestand vorhanden sind. |
| | Bedarfe | Der Bedarf beschreibt die Menge von z. B. einem Material, welches zu einem gewissen Zeitpunkt in einem Werk benötigt wird. |
| Auftragsdaten | Auftragsnummer | Die Auftragsnummer ist die Nummer eines Plan- oder Werkauftrags. |
| | Auftragszeiten | Die Auftragszeit ist die vorgegebene Zeit für die Erledigung eines Auftrags durch einen Menschen. |
| | Arbeitsgangzeiten | Die Arbeitsgangzeit ist die Zeit, welche für die Durchführung eines Arbeitsgangs benötigt wird. |
| | Auftragsfortschritt | Der Auftragsfortschritt beschreibt den aktuellen Erfüllungsgrad eines in Bearbeitung stehenden Auftrags. |
| | Bearbeitungszeit | Die Bearbeitungszeit beschreibt die benötigte Zeit für die reine Bearbeitung eines Auftrags. |
| | Transportzeit | Die Transportzeit ist die Zeit, welche benötigt wird, um Material von einem Arbeitsplatz zu einem anderen zu transportieren. |
| | Liegezeit | Die Liegezeit ist die Zeit zwischen dem Ende der Durchführungszeit und dem Beginn des Transports. |

| | | |
|---|---|---|
| | Rüstzeit | Die Rüstzeit ist die Zeit, welche benötigt wird, um vorbereitende Maßnahmen zur Durchführung von Vorgängen an einem Arbeitsplatz zu treffen. |
| | Durchlaufzeit | Die Durchlaufzeit ist die Zeitspanne zwischen dem Start der ersten Aktivität eines Auftrags und dem Ende der letzten Aktivität des Auftrags. |
| | produzierte Menge | Die produzierte Menge eines Produkts ist die Menge, welche in einem bestimmten Zeitraum produziert wurde. |
| | Ausschuss | Der Ausschuss ist der Anteil der produzierten Menge, die nicht den vorgegebenen Qualitätsanforderungen entspricht. |
| | Störungen | Störungen beschreiben Ereignisse, die nicht zum Standardablauf eines Vorgangs gehören und dessen Durchführung beeinträchtigen. |
| | Maschinenlaufzeit | Die Maschinenlaufzeit ist die Zeit, welche eine Maschine grundsätzlich bei Annahme eines ständigen Betriebs laufen könnte, abzüglich Stillstandszeiten. |

## 8.2 Datenquellen

| Datenquelle | Beschreibung | Beispiel |
|---|---|---|
| Maschinendatenerfassung (MDE)<br><br>(s. Kletti 2015; Schuh et al. 2017) | Die MDE ist ein System zur automatisierten Dokumentation von im Rahmen des Produktionsprozesses unmittelbar an Maschinen und Anlagen entstehenden Informationen. | Speicherprogrammierbare Steuerungen (SPS), Schnittstellen zu Modulen eines Bussystems, mobile Erfassungsgeräte |
| Informationssysteme<br><br>(s. Kletti 2015; Schuh et al. 2017) | Informationssysteme dienen als Drehscheibe des unternehmensinternen Datentransfers, in dem relevante Daten zentral erfasst und verarbeitet werden. | ERP(Enterprise-Resource-Planning)-System: dient der Unterstützung, Bündelung und Steuerung aller notwendigen Geschäftsprozesse innerhalb eines Unternehmens |
| Intelligente Sensorik<br><br>(s. Beyerer et al 2016) | Intelligente Sensorik ist ein Oberbegriff für prozessnahe Erfassung und Verarbeitung relevanter Prozess- und Produktinformationen in der Maschine, welche Regelungssystemen zur Verfügung gestellt werden können. | Erfassung von Umgebungsbedingungen (Temperater, Luftfeuchtigkeit, …), Einsatz als optisches Messgerät zur Erkennung von Produktionsfehlern entlang der Fertigungsstraße |
| Real-Time Location Systems (RTLS)<br><br>(s. Gladysz u. Santarek 2017) | RTLS ermöglicht automatische Lokalisierung von Objekten in Echtzeit. | vereinfachtes Tracken und Navigieren von Personen und Objekten |
| Auto-ID<br><br>(s. Oner et al. 2017) | Auto-ID ermöglicht automatische Identifikation von Objekten und Erfassung ihrer Daten. | Barcodes und Magnetstreifen, Einsatzmöglichkeiten z. B. in der Lagerverwaltung oder Betriebsdatenerfassung |

| | | |
|---|---|---|
| Lesegeräte<br><br>(s. Kletti 2015) | Lesegeräte sind Geräte zum Auslesen codierter Daten. | Barcode-Scanner |
| Buzzer<br><br>(s. Kletti 2015) | Der Buzzer erzeugt bei manueller Betätigung eine systemseitige Rückmeldung. | Erfassung der Arbeitszeiten: Beginn und Ende der Arbeitszeit werden durch Betätigen des Buzzers vermerkt |
| Terminal zur Personalzeiterfassung (PZE)<br><br>(s. Kletti 2015) | Das PZE dient der Erfassung der täglichen Arbeitszeit der Mitarbeiter, z. B. über das Scannen von Stempelkarten. | |
| Human-Machine-Interface (HMI)<br><br>(s. Kletti 2015) | Das HMI (Mensch-Maschine-Schnittstelle) ist eine Schnittstelle, welche Benutzern die Kommunikation mit Maschinen und Systemen ermöglicht. | Anschließen von Bildschirmen und Dashboards an Anlagenkomponenten zur Visualisierung von Daten und Steuerung von Systemen durch den Anwendenden |
| Waagen<br><br>(s. Kletti 2015) | Die Waage ist ein Messgerät zur Bestimmung der Masse eines Objekts. | |
| Begleitpapiere<br><br>(s. KlettI 2015) | Begleitpapiere sind schriftliche Unterlagen, welche die Ware beim Transport begleiten. | Lieferscheine, Kanbankarten. |
| Mechanische Zählwerke | Das mechanische Zählwerk ist ein mechanisches Bauteil, zur Erfassung von z. B. Stückzahlen und Durchflussmengen in der Produktion über entsprechende Sensoren. | |
| Mobile Applications | Mobile-Applications umfassen Softwares für Mobilgeräte. | Softwares zur Anlagenüberwachung auf Mobilgeräten |
| Externe Internetquellen | Als externe Internetquellen werden Internetquellen bezeichnet, welche nicht dem eigenen Unternehmen zuzuschreiben sind. | Informationsbeschaffung von Marktforschungsinstituten oder Wirtschaftsverbänden |
| Terminal zur Betriebsdatenerfassung (BDE)<br><br>(s. Roschmann 1991) | BDE-Terminals ermöglichen die manuelle Eingabe von Betriebsdaten. | Staplerterminals in der Logistik, Visualisierungseinheit in der Produktion. |
| Dokumentation | Die Dokumentation von Daten dient der Beurteilung des Analysepotenzials der Messwerte (z. B. von Maschinendaten) für Dritte. | |
| Fabrikkalender | Der Fabrikkalender ist ein Kalender, in welchem die Arbeitstage fortlaufend nummeriert sind. | Beispiel: Montag bis Freitag sind Arbeitstage. Samstag, Sonntag und Feiertage sind arbeitsfreie Tage. |
| Offline-Datenbank | Eine Offline-Datenbank ist ein elektronisches Datenverwaltungssystem, auf welches auch ohne bestehende Verbindung zum Inter- oder Intranet zugegriffen werden kann. | |

## 8.3 Datenqualitätsmerkmale

| Datenqualitäts-merkmal | Beschreibung (s. Rohweder et al. 2015, S. 27-43; s. Apel et al. 2009, S. 22 f.) |
|---|---|
| Glaubwürdigkeit | Die Glaubwürdigkeit von Daten wird maßgeblich durch ihre Aufbereitung beeinflusst. Daten werden als glaubwürdig bezeichnet, wenn Zertifikate oder ihre Aufmachung einen hohen Qualitätsstandard suggerieren und die Informationsgewinnung und -verbreitung mit einem angemessenen Aufwand betrieben wurden. |
| Genauigkeit | Daten werden als genau bezeichnet, sofern sie in Abhängigkeit des jeweiligen Anwendungsfalls als korrekt und zuverlässig angesehen werden können. Beispielsweise gibt es eine Küchenwaage, die Werte bis auf drei Nachkommastellen genau angibt. |
| Objektivität | Daten werden als objektiv bezeichnet, sofern sie sachlich und wertfrei, also ohne subjektiven Einfluss sind. So ist beispielsweise eine Einschätzung, wie sicher bzw. unsicher ein Land ist, nur dann objektiv, wenn diese Einschätzung durch einen unabhängigen Sachverständigen anhand festgelegter Kriterien getroffen wird. |
| Reputation | Daten besitzen eine gute Reputation bzw. ein hohes Ansehen, wenn die Datenquelle, das Transportmedium und das verarbeitende System erfahrungsgemäß eine hohe Vertrauenswürdigkeit besitzen und eine hohe Güte suggerieren. Daten werden zum Beispiel als qualitativ hochwertig und präzise angesehen, wenn sie von einer Abteilung bereitgestellt werden, welche in der Vergangenheit keine Probleme bei der Datenbereitstellung hatte. |
| Mehrwert | Daten erzielen einen Mehrwert bzw. sind wertschöpfend, wenn ihre Nutzung zu einer quantifizierbaren Steigerung einer monetären Zielfunktion befähigt. Das Gesprächsprotokoll zu einer Reklamation ist beispielsweise wertschöpfend, sofern durch dessen Auswertung ein/e Kund*in gehalten werden kann. |
| Relevanz | Daten werden als relevant bezeichnet, wenn sich der aus ihnen extrahierbare Informationsgehalt in einem definierten Anwendungsfall mit dem Informationsbedarf einer Anfrage deckt. Eine hohe Relevanz hat zum Beispiel die sekundengenaue Angabe des Startzeitpunkts einer Rakete im Gegensatz zu der sekundengenauen Angabe des Anlieferungszeitpunkts von Rohstoffen. |
| Aktualität | Die Aktualität eines Datensatzes beschreibt die Fähigkeit, bei Ände-rungen in der realen Welt zeitnah die entsprechenden Daten anzupassen. Aktualität ist somit die Resistenz eines Datensatzes gegenüber Fehlern aufgrund der zeitlichen Änderung der realen Welt. Ein Beispiel für einen aktuellen Datensatz ist eine Liste von Kundenadressen, welche mit den derzeitigen Standorten der Kunden übereinstimmen. |
| Vollständigkeit | Daten werden als vollständig bezeichnet, wenn zu einem festgelegten Zeitpunkt alle für einen Prozessschritt benötigten Daten zur Verfügung stehen. Ein Beispiel für einen vollständigen Datensatz ist eine Liste von Kundenadressen, welche ausnahmslos von allen Kunden die Adressen beinhaltet. |
| Datenmenge | Ein Datensatz besitzt einen angemessenen Umfang bzw. die Datenmenge ist ausreichend groß, sofern die Menge der verfügbaren Daten einerseits den gestellten Anforderungen genügt und andererseits nicht überflüssig groß ist. Ein Beispiel für einen ausreichenden Umfang ist eine Datenbank, in welcher genügend Kundentransaktionen gespeichert sind, um sinnvolle Rückschlüsse auf das Kaufverhalten zu ziehen. |

| | |
|---|---|
| Interpretierbarkeit | Die Interpretierbarkeit von Daten beschreibt das Ausmaß, in dem Daten in einer verständlichen Sprache vorliegen und verwendete Maßeinheiten bzw. Definitionen verständlich sind. Beispielsweise ist das Gewicht eines Gegenstands nur sinnvoll interpretierbar, sofern bekannt ist, in welcher Einheit dieses erhoben wurde. |
| Verständlichkeit | Daten werden als verständlich bezeichnet, wenn sie unmittelbar von den Anwendenden verstanden werden und für deren Zwecke einsetzbar sind. Ein Beispiel für einen verständlichen Datensatz ist eine Liste von Kundenadressen, welche die Postleitzahlen und Orte sowie die Straßen und Hausnummern der Kunden beinhaltet. |
| Einheitlichkeit | Daten werden als einheitlich bezeichnet, wenn sie unabhängig vom Zeitpunkt und datenerfassenden Personen im selben Format, im selben Layout und mit derselben Wertemenge beschrieben werden. Einheitlichkeit ist beispielsweise bei der Beschreibung des Geschlechts mit der Wertemenge {m, w, d} gegeben. |
| Übersichtlichkeit | Die Darstellung von Daten wird als übersichtlich bezeichnet, wenn genau die benötigten Informationen in einem passenden und leicht erfassbaren Format dargestellt sind. Ein Beispiel für eine übersichtliche Darstellung eines Datensatzes sind Candle-Stick-Charts zur Darstellung von Kursentwicklungen. |
| Erreichbarkeit | Daten werden als erreichbar bezeichnet, sofern diese anhand einfacher Verfahren und auf direktem Weg abrufbar sind. Ein Beispiel für einfach erreichbare Daten ist der mögliche Aufruf von Kundenstammdaten in einem System anhand der Kundennummer. |
| Zugriffssicherheit | Die Zugriffssicherheit von Daten ist das Ausmaß, in dem der Zugang zu Daten eingeschränkt und kontrolliert werden kann. Beispielsweise sollten Anwender in Unternehmen nur auf die Daten und Anwendungen zugreifen können, für die sie auch eine Berechtigung besitzen. |

## 8.4 Morphologische Betrachtung von Datenquellen

| Datenmerkmal | Beschreibung | Differenzierungsmöglichkeiten |
|---|---|---|
| Herkunft | Die Herkunft beschreibt den Ursprungsort der Datenquellen. | Man unterscheidet zwischen innerhalb eines Unternehmens bezogenen Daten (Herkunft intern) und außerhalb eines Unternehmens bezogenen Daten (Herkunft extern). |
| Metadaten<br><br>(s. Apel et al 2009, S. 208 ff.) | Metadaten enthalten Informationen über den Aufbau, die Struktur und den Inhalt der betrachteten Daten. | Metadaten sind entweder vorhanden, teilweise vorhanden (nur für bestimmte Attribute oder Datensätze) oder nicht vorhanden. |
| Art der Erfassung<br><br>(s. VDI 2016, S. 42) | Die Art der Erfassung kann anhand ihres Automatisierungsgrades differenziert werden. | Man unterscheidet zwischen automatischer, halbautomatischer und manueller Datenerfassung. |
| Erfassungsauslösung<br><br>(s. Reinhardt 1996) | Die Erfassungsauslösung ist die Ursache der Datenerfassung. | Die Erfassungsauslösung erfolgt zeitlich kontinuierlich, in festgelegten Intervallen oder durch das Eintreffen eines definierten Ereignisses. |

| | | |
|---|---|---|
| Archivierungszeitraum | Der Archivierungszeitraum ist der Zeitraum, über welchen erfasste Daten gespeichert werden. | Der Archivierungszeitraum ist variabel einstellbar und schwankt zwischen Zeiträumen, welche kürzer als ein Tag und länger als ein Monat sind. |
| Aktualisierungsrate | Die Aktualisierungsrate ist die Zeitspanne zwischen der Speicherung eines Datensatzes und dessen Aktualisierung. | Die Aktualisierungsrate ist variabel einstellbar und kann Werte zwischen weniger als einer Sekunde und länger als einer Stunde annehmen. |
| Schnittstellen | Eine Schnittstelle ist eine Möglichkeit, die von einer Datenquelle generierten Daten abzugreifen. | Es wird differenziert zwischen<br>Maschinenschnittstellen (z. B. einem USB-Anschluss),<br>Softwareschnittstellen (z. B. einer Exportfunktion),<br>Mensch-Maschine-Schnittstellen (z. B. einem Bedienterminal) und<br>Mensch-Mensch-Schnittstellen (Medien der zwischenmenschlichen Kommunikation). |
| Datentyp<br><br>(s. Lange u. Stegemann 1985) | Der Datentyp beschreibt eine Menge von Datenobjekten, welche die gleiche Struktur haben und mit denen die gleichen Operationen ausgeführt werden können. | Es wird differenziert zwischen den Datentypen<br>Integer (ganze Zahlen)<br>Float (rationale Zahlen),<br>Boolean (Wahrheitswerte) und<br>String (Zeichenketten). |
| Datentyp<br><br>(s. Schramm 2008) | Die Datenstruktur ist eine Möglichkeit der Datenspeicherung und Organisation. | Strukturierte Daten besitzen ein festgelegtes Schema und werden als Attributwerte zu zugehörigen Attributen definiert. |
| | | Semistrukturierte Daten tragen einen Teil der Strukturinformation in sich. Ein Beispiel ist ein Datensatz, der in der Datenmodellierungssprache *Extensible Markup Language (XML)* repräsentiert wird. Unstrukturierte Daten besitzen keine feste Struktur. Ein Beispiel ist der Text einer E-Mail. |
| Skalenniveau<br><br>(s. Niederée u. Mausfeld 1996) | Der Begriff „Skalenniveau" kommt aus der Statistik und beschreibt die Skalierbarkeit einer Messung. Diese ist abhängig von möglichen Ausprägungen der Messwerte eines Datensatzes. | Anhand einer *Nominalskala* lassen sich Objekte unterscheiden, jedoch nicht in einer Rangfolge einordnen (z. B. Nachnamen, Autokennzeichen). |
| | | Anhand einer *Ordinalskala* lassen sich Objekte unterscheiden und in einer Rangfolge einordnen (z. B. Schulnoten, Dienstränge). |
| | | Anhand einer *Intervallskala* lassen sich Objekte unterscheiden und in einer Rangfolge anordnen. Abstände zwischen einzelnen Objekten sind quantifizierbar. Nullpunkte sind willkürlich festgelegt (z. B. Temperatur in Grad Celsius, Jahreszahlen). |
| | | Anhand einer *Verhältnisskala* lassen sich Objekte unterscheiden und in einer Rangfolge anordnen. Abstände zwischen den einzelnen Objekten sind quantifizierbar. Durch die Existenz eines natürlichen Nullpunktes lassen sich Verhältnisse zwischen den Größen zweier Objekte bestimmen (z. B. Temperatur in Kelvin, Lebensalter eines Menschen). |

| | | |
|---|---|---|
| Auflösung<br><br>(s. Weichert u. Wülker 2010, S. 11) | Die Auflösung ist der kleinstmögliche Abstand zwischen einem Messwert und dem nächsthöheren Messwert. Dieser Abstand wird bezogen auf die Gesamtbreite des Quantisierungsbereichs. | Die Auflösung ist abhängig von dem gewählten Messinstrument. Sie kann zwischen Werten unterhalb von 0,1 % und oberhalb von 10 % variieren.<br><br>Beispiel: Eine Waage kann Objekte mit einem Maximalgewicht von bis zu 100 kg wiegen. Das Ergebnis eines Wiegevorgangs kann bis auf die erste Nachkommastelle genau angegeben werden. Der Quantisierungsbereich beträgt 0 kg – 100 kg, die Auflösung dementsprechend 0,1/100 = 0,1 %. |
| Störanfälligkeit | Die Störanfälligkeit ist ein Maß für die Robustheit der Datenerfassung gegen äußere Einflüsse und die Resistenz gegenüber Datenfehlern. | Differenziert wird zwischen einer geringen, mittleren und hohen Störanfälligkeit. Beispielsweise ist die Störanfälligkeit gering, sofern Messwerte kaum von Umwelteinflüssen abhängen. Eine hohe Störanfälligkeit liegt z. B. vor, wenn Messwerte durch manuelle Erfassung abhängig von den Anwendenden sind. |

## 8.5 Methoden der Datenfusion

### 8.5.1 Entscheidungsregeln

Entscheidungsregeln entsprechen konkreten Heuristiken, die im Falle konfligierender Daten spezifische Handlungsanweisungen geben[37].

[37] s. Bleiholder u. Naumann 2006, S. 3 ff.; 2008, S. 9

| Klassifikation | Strategie | Beschreibung |
|---|---|---|
| Konflikte vermeiden | Pass it on | Nach der Entscheidungsregel „Pass it on“ werden konfligierende Attributwerte gemeinsam übernommen. Es wird den Anwendenden des fusionierten Datenbestands überlassen, wie die auftretenden Datenkonflikte behandelt werden. |
| | Consider all Possibilities | Nach der Entscheidungsregel „Consider all Possibilities“ werden alle möglichen Kombinationen der konfligierenden Attributwerte erzeugt und in den fusionierten Datenbestand übernommen. Problematisch ist hier die Erzeugung von Attributwerten, die nicht in den Ausgangsbeständen enthalten sind. |

| | | |
|---|---|---|
| Konflikte ignorieren | Take the Information | Nach der Entscheidungsregel „Take the Information“ sind bei einer Subsumption zweier Dubletten die Attributwerte zu wählen, die keine Nullwerte sind. Diese Strategie eignet sich somit nur zur Anwendung bei Unsicherheiten und nicht bei Widersprüchen. |
| | No Gossiping | Nach der Entscheidungsregel „No Gossiping“ werden nur konsistente Datensätze in den fusionierten Datenbestand übernommen. Dubletten, welche Unsicherheiten oder Widersprüche enthalten, werden gelöscht. |
| | Trust your Friends | Nach der Entscheidungsregel „Trust your Friends“ werden Metadaten verwendet, um Präferenzen bezüglich der Datenquellen zu bilden. Anhand dieser Präferenzen werden im Konfliktfall die Attributwerte aus der Datenquelle mit der höheren Präferenz übernommen. Beispielsweise können sich bestimmte Datenquellen durch eine erfahrungemäß hohe Verlässlichkeit oder hohe Aktualität auszeichnen. |
| Konflikte auflösen | Cry with the Wolves | Nach der Entscheidungsregel „Cry with the Wolves“ wird derjenige Attributwert ausgewählt, welcher in den betrachteten Dubletten am häufigsten vertreten ist. |
| | Roll the Dice | Nach der Entscheidungsregel „Roll the Dice“ wird bei zwei miteinander in Konflikt stehenden Attributwerten einer zufällig ausgewählt. Klarer Vorteil hier ist der geringe Anwendungsaufwand, die Auswahl der Attributwerte erfolgt jedoch nahezu willkürlich. |
| | Meet in the Middle | Nach der Entscheidungsregel „Meet in the Middle“ wird ein neuer Wert aus den konfligierenden Attributwerten erzeugt, z. B. der Mittelwert, welcher die Gesamtdistanz zu den Ausgangswerten minimiert. |
| | Keep up to Date | Nach der Entscheidungsregel „Keep up to the Date“ wird der Attributwert gewählt, welcher auf Basis des Zeitstempels des zugehörigen Datensatzes der aktuelle ist. |
| | Ignorance is Bliss | Nach der Entscheidungsregel „Ignorance is Bliss“ wird der Attributwert nur aus den erfahrungsgemäß validen Attributwerten gewählt. Unplausible Werte werden ignoriert. |
| | Oops, I did it again | Nach der Entscheidungsregel „Oops, I did it again“ wird derjenige Attributwert gewählt, für den man sich in einer ähnlichen Situation in der Vergangenheit bereits erfolgreich entschieden hatte. |
| | Better bend than break | Nach der Entscheidungsregel „Better bend than break“ wird der letzte gemeinsame Vorfahre von zwei miteinander in Konflikt stehenden Attributwerten auf Basis der Taxonomie oder Syntax ermittelt. Beispielsweise werden die Klassifikationen *Audi A8* und *Mercedes AMG GT* zu der Klassifikation als Sportwagen aggregiert. |

### 8.5.2 Methoden der klassischen Statistik

Die Methoden der klassischen Statistik basieren auf einer frequentischen Interpretation der Wahrscheinlichkeiten, bei welcher diese als Grenzwerte relativer Häufigkeiten betrachtet werden. Daraus ergibt sich für den Anwendungsfall ein wahrscheinlichkeitstheoretisches Modell, das Wahrscheinlichkeiten als Grenzwert beobachtbarer Häufigkeiten definiert. Damit werden die beobachteten, zu fusionierenden Daten als empirische Repräsentation einer Zufallsvariablen Y betrachtet. Die Wahrscheinlichkeitsverteilung dieser Zufallsvariablen ist abhängig von einer zugehörigen tatsächlichen, jedoch unbekannten Messgröße $\theta$. Die Schätzung dieser Messgröße anhand der vorliegenden Daten ist das Ergebnis der Datenfusion[38].

In der folgenden Tabelle sind die zwei gängigsten Methoden der klassischen Statistik aufgeführt:

| | |
|---|---|
| Maximum-Likelihood-Methode | Die Maximum-Likelihood-Methode ermittelt denjenigen Schätzer, der als Parameter der Wahrscheinlichkeitsverteilung der Zufallsvariablen die Realisierung dieser Zufallsvariablen gemäß den beobachteten Daten am wahrscheinlichsten macht (s. Klaus 1999, S. 89 – 103; vgl. Utschick u. Dietrich 2006). |
| Ansatz der kleinsten Fehlerquadrate | Mit dem Ansatz der kleinsten Fehlerquadrate wird derjenige Schätzer ermittelt, mit dem das zugrundeliegende Modell die beobachteten Repräsentationen der zu untersuchenden Zufallsvariablen am genauesten beschreibt. Dazu wird die Summe der quadrierten Differenzen zwischen den einzelnen Repräsentationen und der Modellprognose in Abhängigkeit des zu schätzenden Parameters gebildet. Für die Optimalität des Vorhersagemodells wird derjenige Parameter $\theta'$ gewählt, der die Summe der Fehlerquadrate und somit den Abstand zwischen den prognostizierten und beobachteten Werten minimiert (s. Myung 2003, S. 95). |

### 8.5.3 Bayes'sche Inferenz

Im Gegensatz zu der klassischen Statistik wird der *Bayes'schen Statistik* die Interpretation der Wahrscheinlichkeiten als *Degree of Belief (DoB)* zugrunde gelegt. Der DoB repräsentiert für ein Ereignis den Grad der Überzeugung bezüglich des Eintretens des Ereignisses auf Basis der vorliegenden Daten. Aufgrund dieser Eigenschaft ist der DoB grundsätzlich von der Subjektivität der definierenden Person abhängig, da unterschiedliche Individuen das Eintreten desselben Ereignisses bei gleichem Informationsstand möglicherweise unterschiedlich bewerten[39].

| | |
|---|---|
| Bayes'sche Inferenz | Ermittlung derjenigen Hypothese aus einem Kreis definierter Einzelhypothesen, welche auf Basis ihrer A-posteriori-Wahrscheinlichkeiten und unter Beobachtung eines Ereignisses Y am wahrscheinlichsten ist. Zentrale Werkzeuge sind das Bayes-Theorem und der Satz der totalen Wahrscheinlichkeit (s. Ruser u. Puente León 2007, S. 98). |

[38] s. Beyerer et al. 2006, S. 25; s. Ruser u. Puente León 2007, S. 98

[39] s. Beyerer et al. 2006, S. 25

### 8.5.4 Dempster-Shafer Methode

Die Dempster-Shafer-Methode erweitert die Wahrscheinlichkeit um das zweidimensionale Maß der Evidenz. Diese setzt sich zusammen aus dem DoB und der Plausibilität, dem Maß für die maximale Möglichkeit der Korrektheit einer Hypothese.

| | |
|---|---|
| Dempster-Shafer-Methode | Ermittlung derjenigen Hypothese aus einem Kreis definierter Einzelhypothesen, welche das kleinste Evidenzintervall besitzt. Dieses wird durch Werte einer Vertrauens- bzw. Plausibilitätsfunktion begrenzt, welche für jede Hypothese den DoB und die Plausibilität berechnen. Je enger das Intervall ist, desto vertrauenswürdiger ist die Hypothese (s. Ruser u. Puente León 2007, S. 99; s. Dietmayer 2006, S. 39 ff.). |

### 8.5.5 Fuzzy-Logik

Die Fuzzy-Logik wird zur Modellierung von Ungewissheit oder Vagheit verwendet. Dabei ist zwischen den Begriffen der Unsicherheit und der Ungewissheit zu unterscheiden. Während beispielsweise Unsicherheit darüber besteht, ob am nächsten Tag gutes Wetter sein wird, bezeichnet Ungewissheit die unscharfe (engl. *fuzzy*) Definition, was genau gutes Wetter bedeutet. Falls z. B. für gutes Wetter eine Temperatur von mindestens 25 °C notwendige Bedingung ist, erscheint eine Klassifikation als nicht gutes Wetter bei einer Temperatur von 24,99 °C fragwürdig.

| | |
|---|---|
| Fuzzy-Logik | Die Fuzzy-Logik bietet eine kontinuierlich abgestufte statt absolute Zuordnung von Objekten zu bestimmten Klassen. So wird eine Menge nicht durch die in ihr enthaltenden Elemente definiert, sondern durch den Grad ihrer Zugehörigkeit zu dieser Menge. Im Gegensatz zur klassi-schen Mengenlehre, nach der ein Element entweder Teil einer Menge ist oder nicht, ist damit das Maß der Zugehörigkeit eines Elements zu einer Menge explizit formulierbar (s. Ruser u. Puente León 2006, S. 11; 2007, S. 99; s. Lehmann et al. 1992, S. 1 f.). |

### 8.5.6 Künstliche Neuronale Netze (KNN)

Künstliche Neuronale Netze sind analog zu biologischen neuronalen Netzten aufgebaut. Informationen werden von Neuronen verarbeitet und über gewichtete Verbindungen weitergegeben. Es wird differenziert zwischen einer Eingabe-, Ausgabe- und versteckten Schicht von Neuronen.

| | |
|---|---|
| Künstliche Neuronale Netze | Eingabeneuronen erhalten eine direkte Eingabe aus der Umgebung. Die Informationsverarbeitung findet in den versteckten Schichten statt und erfolgt nach der Funktionsweise eines Schwellwertelements. Zunächst werden alle gewichteten Eingänge summiert. Mit einer Akti-vierungsfunktion werden die Eingangsgrößen bewertet. Wird ein Neuron durch genügend positive Einträge erregt, ohne gleichzeitig von zu vielen negativen Eingängen gehemmt zu werden, sendet es ein Ausgangssignal. Auf diese Weise wird erst ab dem Erreichen eines Schwellwertes ein signifikantes Ausgangssignal gesendet. Dieses Ausgangssignal wird den Kantengewichten entsprechend gewichtet und ist in einem Neuronalen Netz neues Eingangssignal für Neuronen auf der nachgelagerten Ebene. In der Ausgabeschicht können die Anwendenden die Zielinformationen dann auswerten (s. Fritsch u. Finke 2012, S. 307; vgl. Lawrence et al. 2012). |

### 8.5.7 Relationale Operatoren

Relationale Operatoren kombinieren verschiedene Datenquellen, die in Form von Tabellen vorliegen. Sie sind daher nicht dazu geeignet, konfligierende Datensätze zu fusionieren und sind somit im Kontext der Datenfusion lediglich von geringem Nutzen.

| | |
|---|---|
| Relationale Operatoren | Bei der Kombination verschiedener Datenquellen besteht ein Trade-off zwischen einer simultanen Erhöhung der Vollständigkeit auf der einen und der Exaktheit und Prägnanz auf der anderen Seite. Die Vollständigkeit wird durch Berücksichtigung mehrerer Datenquellen, Objekte oder Attribute erhöht. Die Prägnanz jedoch wird durch die Entfernung jener erhöht. Die relationalen Operatoren können in Union- und Join-Operatoren unterteilt werden (s. Bleiholder u. Naumann 2008, S. 5). |

# 9 Checkliste zur Anwendung der Datenfusion im Kontext betrieblicher Rückmeldedaten

Die hier angefügte Checkliste dient als Hilfestellung zur Umsetzung des in diesem Leitfaden erarbeiteten Konzepts zur Anwendung der Datenfusion bei der Erfassung und Speicherung betrieblicher Rückmeldedaten. Den Anwendenden ist es möglich, dass in diesem Leitfaden vermittelte Konzept in sieben Schritten für ihre individuellen Anwendungsfälle durchzuführen.

**Checkliste: Datenfusion in 7 Schritten**

Startdatum:

Durchführende Person(en)/ Abteilung:

| | Arbeitsschritt | Kommentar | Erledigt |
|---|---|---|---|
| | Definition des Anwendungsfalls | | |
| 1 | Ermittlung der Informationsbedarfe des PPS-Systems | | |
| 2 | Feststellen der Informationsverfügbarkeit auf Basis der vorhandenen Datenquellen | | |
| | Bestimmung der zu fusionierenden Datenquellen | | |
| 3 | Zuordnung der vorhandenen Datenquellen zu bestehenden Informationsbedarfen | | |
| 4 | Bewertung der Datenqualität der Datenquellen | | |
| | Auswahl geeigneter Methoden der Datenfusion | | |
| 5 | Morphologische Einordnung der vorhandenen Datenquellen | | |
| 6 | Ableitung prozesstypischer Fehler | | |
| 7 | Zuordnung von Methoden der Datenfusion zu prozesstypischen Fehlern | | |
| | Abschluss | | |

# 10 Literaturverzeichnis


Apel, D.; Behme, W.; Eberlein, R.; Merighi, C.: Datenqualität erfolgreich steuern. Praxislösungen für Business-Intelligence-Projekte. Hanser, München [u. a.] 2009.

Becker, W.; Ulrich, P.; Botzkowski, T.: Industrie 4.0 im Mittelstand. Best Practices und Implikationen für KMU. Springer Gabler, Wiesbaden 2017.

Beyerer, J.; Sander, J.; Werling; S.: Fusion heterogener Informationsquellen. In: Informationsfusion in der Mess- und Sensortechnik. Hrsg.: J. Beyerle. Universitätsverlag, Karlsruhe 2006, S. 21 – 37.

Beyerer, J.; Puento León, F.; Frese, C.: Automatische Sichtprüfung. Grundlagen, Methoden und Praxis der Bildgewinnung und Bildauswertung. 2., erw. u. verb. Auflage. Springer Vieweg, Berlin [u. a.] 2016.

Bleiholder, J.; Naumann, F.: Conflict Handling Strategies in an Integrated Information System. Institut für Informatik der Humboldt-Universität zu Berlin, April 2006. https://www.researchgate.net/publication/238121998_Conflict_Handling_Strategies_in_an_Integrated_Information_System#read (Link zuletzt geprüft: 20.05.2021)

Bleiholder, J.; Naumann, F.: Data fusion. In: ACM Computing Surveys 41 (2008) 1, S. 1 – 41.

Bleiholder, J.; Naumann, F.: Kurz erklärt: Datenfusion. In: Datenbank-Spektrum 11 (2011) 1, S. 59 – 61.

Bleiholder, J.; Schmid, J.: Datenintegration und Deduplizierung. In: Daten- und Informationsqualität. Auf dem Weg zur Information Excellence. Hrsg.: K. Hildebrand; M. Gebauer; H. Hinrichs; M. Mielke. 3., erw. Auflage. IT. Springer Vieweg, Wiesbaden 2015, S. 121 – 140.

Bley, A.; Vogt, G.; Holstein, M.; Niegsch, C.; Morales, L. M.: Mittelstand im Mittelpunkt. Eine Publikation von BVR und DZ BANK AG. Ausgabe Früjahr 2019. Volkswirtschaft; Nr. 10. 11.06.2019. https://www.bvr.de/p.nsf/0/494DDCB46E8A6E01C125841D0041BF72/%24FILE/Mittelstand%20im%20Mittelpunkt%20Fr%C3%BChjahr%202019.pdf (Link zuletzt geprüft: 20.05.2021)

Dienes, C.; Pahnke, A.; Wolter, H.-J.: Investitionsverhalten von kleinen und mittleren Unternehmen. IfM-Materialien; Nr. 268. Institut für Mittelstandsforschung (IfM), Bonn 2018. https://www.ifm-bonn.org/fileadmin/data/redaktion/publikationen/ifm_materialien/dokumente/IfM-Materialien-268_2018.pdf (Link zuletzt geprüft: 20.05.2021)

Dietmayer, K.: Evidenztheorie: Ein Vergleich zwischen Bayesund Dempster-Shafer-Methoden. In: Informationsfusion in der Mess- und Sensortechnik. Hrsg.: J. Beyerer. Universitätsverlag, Karlsruhe 2006, S. 39 – 49.

DURRANT-WHYTE, H.; HENDERSON T. C.: Multi Sensor Data Fusion. In: Springer Handbook of Robotics. Hrsg. B. Siciliano; O. Khatib. Springer, Berlin, Heidelberg 2008, S. 585 – 610.

Farkisch, K.: Data-Warehouse-Systeme kompakt. Aufbau, Architektur, Grundfunktionen. Springer, Berlin [u. a.] 2011.

Fauvel, M.; Chanussot, J.; Benediktsson, J. A.: Decision Fusion for the Classification of Urban Remote Sensing Images. In: IEEE Transactions on Geoscience and Remote Sensing 44 (2006) 10, S. 2828 – 2838.

Fritsch, J.; Finke, M.: Applying Divide and Conquer to Large Scale Pattern Recognition Tasks. In: Neural Networks: Tricks of the Trade. Theoretical Computer Science and General Issues; Bd. 7700. Hrsg.: G. Montavon; G. B. Orr; K.-R. Müller. Springer, Berlin [u. a.] 2012, S. 311 – 338.

Gladysz, B.; Santarek, K.: An Approach to RTLS Selection. In: 24th International Conference on Production research (ICPR 2017). Posnan, Poland, July 30 – August 3, 2017. DEStech Publications, Lancaster 2017, S. 13 – 18. https://www.researchgate.net/publication/323177691_AN_APPROACH_TO_RTLS_SELECTION#read (Link zuletzt geprüft: 20.05.2021) Ist es das?

Jirak, J. M.; Krenz, K.; Reiss, M.; Trabs, M.: Methoden der Statistik. Version vom 10. Oktober 2018. https://www.math.hu-berlin.de/~mreiss/MethodenDerStatistik_20181010.pdf 8Link zuletzt geprüft: 20.05.2021)

Klaus, F.: Einführung Techniken und Methoden der Multisensor-Datenfusion. Universität Siegen, Siegen 1999. https://dspace.ub.uni-siegen.de/bitstream/ubsi/57/1/klaus.pdf (Link zuletzt geprüft: 20.05.2021)

Kletti, J.: MES – Manufacturing Execution System. 2. Auflage. Springer Vieweg, Berlin [u. a.] 2015.

Kurbel, K. E.: Enterprise Resource Planning and Supply Chain Management. Functions, Business Processes and Software for Manufacturing Companies Springer, Berlin [u. a.] 2013.

Lange, O.; Stegemann, G.: Datenstrukturen und Speichertechniken. Vieweg+Teubner, Wiesbaden 1985.

Lawrence, S.; Burns, I.; Back, A.; Tsoi, A. C.; Giles, C. L.: Neural Network Classification and Prior Class Probabilities. In: Neural Networks: Tricks of the Trade. Theoretical Computer Science and General Issues; 7700. Hrsg.: G. Montavon; G. B. Orr; K.-R. Müller. Springer, Berlin [u. a.] 2012, S. 295 – 309.

Lehmann, I.; Weber, R.; Zimmermann, H.-J.: Fuzzy Set Theory. Die Theorie der unscharfen Mengen. In: OR Spektrum 14 (1992) 1, S. 1 – 9.

Leser, U.; Naumann, F.: Informationsintegration. Architekturen und Methoden zur Integration verteilter und heterogener Datenquellen. dpunkt-Verl., Heidelberg 2007.

Loos, P.: Grunddatenverwaltung und Betriebsdatenerfassung als Basis der Produktionsplanung und -steuerung. Hrsg.: H. Corsten; B. Friedl. Vahlen, München 1999.

Myung, I. J.: Tutorial on maximum likelihood estimation. In: Journal of Mathematical Psychology 47 (2003) 1, S. 90 – 100.

Nelles, O.: Neuronale Netze: Eine Übersicht. In: Informationsfusion in der Mess- und Sensortechnik. Hrsg.: J. Beyerer. Universitätsverlag, Karlsruhe 2006, S. 93 – 112.

Niederée, R.; Mausfeld, R.: Skalenniveau, Invarianz und „Bedeutsamkeit". In: Handbuch psychologischer Methoden. Hrsg.: E. Erdfelder; R. Mausfeld; T. Meiser; G. Rudinger. [Beltz] Psychologie VerlagsUion (PVU), Weinheim 1996, S. 399 – 410. https://www.researchgate.net/profile/Rainer-Mausfeld/publication/317786336_Skalenniveau_Invarianz_und_Bedeutsamkeit/links/594c0215a6fdcc14c97d8a8e/Skalenniveau-Invarianz-und-Bedeutsamkeit.pdf (Link zuletzt geprüft: 20.05.2021)

Nyhuis, P.; Schmidt, M.; Hübner, M.: Transparenz durch Datenverfügbarkeit als Enabler für eine leistungsfähigere PPS. In: Handbuch Industrie 4.0. Geschäftsmodelle, Prozesse, Technik. Hrsg.: G. Reinhart. Hanser, München [u. a.] 2017, S. 33 – 34.

Oner, M.; Ustundag, A.; Budak, A.: An RFID-based tracking system for denim production processes. In: The International Journal of Advanced Manufacturing Technology 90 (2017) 1-4, S. 591 – 604.

Reinhardt, H.: Automatisierungstechnik. Theoretische und gerätetechnische Grundlagen, SPS. Springer, Berlin [u. a.] 1996.

Rohweder, J. P.; Kasten, G.; Malzahn, D.; Piro, A.; Schmid, J.: Informationsqualität – Definitionen, Dimensionen und Begriffe. In: Daten- und Informationsqualität. Auf dem Weg zur Information Excellence. Hrsg.: K. Hildebrand; M. Gebauer; H. Hinrichs; M. Mielke. 3., erw. Auflage. Springer Vieweg, Wiesbaden [u. a.] 2015, S. 25 – 46.

Rommelfanger, H.: Fuzzy-Logik basierte Verarbeitung von Expertenregeln. In: OR Spektrum 15 (1993) 1, S. 31 – 42.

Roschmann, K.: Betriebsdatenerfassung. In: CIM-Handbuch. Wirtschaftlichkeit durch Integration. Hrsg.: U. W. Geitner. Vieweg+Teubner Verlag, Wiesbaden 1991, S. 89 – 102.

Runkler, T. A.: Data Mining. Methoden und Algorithmen intelligenter Datenanalyse. Vieweg+Teubner, Wiesbaden 2010.

Ruser, H.; Puente León, F.: Methoden der Informationsfusion – Überblick und Taxonomie. In: Informationsfusion in der Mess- und Sensortechnik. Hrsg.: J. Beyerer. Universitätsverlag, Karlsruhe 2006, S. 1 – 20.

Ruser, H.; Puente León, F.: Informationsfusion – Eine Übersicht (Information Fusion – An Overview). In: tm – Technisches Messen 74 (2007) 3, S. 74.

Schramm, M.: Informationsextraktion auf Basis strukturierter Daten. Dresden, Techn. Univ., Dipl.-Arb., 2008. https://www.rn.inf.tu-dresden.de/uploads/Studentische_Arbeiten/Diplomarbeit_Schramm_Marcus_n.pdf (Link zuletzt geprüft: 20.05.2021)

Schuh, G.; Brandenburg, U.; Cuber, S.: Aufgaben. In: Produktionsplanung und -steuerung; Bd. 1: Grundlagen der PPS. Hrsg.: G. Schuh; V. Stich. Springer Vieweg, Berlin [u. a.] 2012, S. 29 – 81.

Schuh, G.; Nyhuis, P.; Reuter, C.; Hauptvogel, A.; Schmitz, S.; Nywlt, J.; Brambring, F.; Schulte, F.; Hansen, J.: Produktionsdaten als Enabler für Industrie 4.0. In: Werkstattstechnik online 105 (2015) 4, S. 200 – 203.

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.

Utschick, W.; Dietrich, F. A.: On estimation of structured covariance matrices. In: Informationsfusion in der Mess- und Sensortechnik. Hrsg.: J. Beyerer. Universitätsverlag, Karlsruhe 2006, S. 51 – 62.

VDI 5600: Fertigungsmanagementsysteme (Manufacturing Execution Systems – MES). VDI-Richtlinie; ICS 35.240.50. VDI – Verein Deutscher Ingenieure (Hrsg.); Düsseldorf, Oktober 2016. https://www.vdi.de/fileadmin/pages/vdi_de/redakteure/richtlinien/inhaltsverzeichnisse/2436698.pdf (Link zuletzt geprüft: 20.05.2021)

Wang, R. Y.; Strong, D. M.: Beyond Accuracy: What Data Quality Means to Data Consumers. In: Journal of Management Information Systems 12 (1996) 4, S. 5 – 33.

Weichert, N.; Wülker, M.: Messtechnik und Messdatenerfassung. 2., aktualis. u, erw. Auflage Oldenbourg, München [u. a.] 2010.

FIR e. V.
an der RWTH Aachen
Campus-Boulevard 55
52074 Aachen

Telefon: +49 241 47705-0
E-Mail: info@fir.rwth-aachen.de
www.fir.rwth-aachen.de