



Stammdatenmanagement Whitepaper

Impressum

Herausgeber:

Univ. Prof. Dr.-Ing. Dipl.-Wirt. Ing. Günther Schuh
Direktor des FIR e. V. an der RWTH Aachen

Prof. Dr.-Ing. Volker Stich
Geschäftsführer des FIR e. V. an der RWTH Aachen

Autoren:

Dipl.-Ing. Steffen Nienke
Bereich Informationsmanagement
Leiter Fachgruppe Informationslogistik
FIR e. V. an der RWTH Aachen

Gregor Josef Fuhs, M.Sc.
Bereich Informationsmanagement
Fachgruppe Informationslogistik
FIR e. V. an der RWTH Aachen

Autoren der 1. Auflage:

Marcel Scheibmayer
Eric Naß
Martin Birkmeier

Bildnachweis:

Fotos: © Fotolia
Diagramme: © FIR e. V. an der RWTH Aachen

Korrektorat:

Simone Suchan M.A.,
FIR e. V. an der RWTH Aachen

Gestaltung, Bildbearbeitung, Satz und Layout:

Birgit Kreitz,
FIR e. V. an der RWTH Aachen

Druck:

Druckservice Zillekens

Lizenzbestimmungen/Copyright

Dieses Werk ist urheberrechtlich geschützt. Die dadurch begründeten Rechte, insbesondere die der Übersetzung, des Nachdrucks, des Vortrags, der Entnahme von Abbildungen und Tabellen, der Funksendung, der Mikroverfilmung oder der Vervielfältigung auf anderen Wegen und der Speicherung in Datenverarbeitungsanlagen, bleiben, auch bei nur auszugsweiser Verwertung, vorbehalten.

Eine Vervielfältigung dieses Werkes oder von Teilen dieses Werkes ist auch im Einzelfall nur in den Grenzen der gesetzlichen Bestimmungen des Urheberrechtsgesetzes der Bundesrepublik Deutschland vom 9. September 1965 in der jeweils gültigen Fassung zulässig. Sie ist grundsätzlich vergütungspflichtig. Zuwiderhandlungen unterliegen den Strafbestimmungen des Urheberrechtsgesetzes.

© 2011
2., neu bearb. Auflage, Januar 2017.

© 2017
FIR e. V. an der RWTH Aachen
Campus-Boulevard 55
52074 Aachen
Tel.: +49 241 47705-0
Fax: +49 241 47705-199
E-Mail: info@fir.rwth-aachen.de
Internet: www.fir.rwth-aachen.de

Inhaltsverzeichnis

1 Management-Summary	5
2 Motivation	7
2.1 Stammdaten und Datenqualität	8
2.2 Auswirkungen und Nutzenaspekte von Datenqualität	10
2.3 Herausforderungen beim Stammdatenmanagement	11
3 Unternehmensweites Stammdatenmanagement	12
3.1 Stammdatenstrategie	12
3.2 Organisatorische Verankerung	12
3.3 IT-Architekturen für das Stammdatenmanagement	15
3.4 Stammdatenstruktur	17
4 Vorgehen bei Stammdatenmanagementprojekten	19
5 Stammdaten und Big Data / Smart Data	20
5.1 Big Data	21
5.2 Smart Data	22
5.3 Smart Information	24
5.4 Beispiele und Anwendungsfälle	24
6 Das FIR als kompetenter Partner in der Praxis	26
7 Literaturverzeichnis	27

Haben Sie Interesse an weiteren Informationen aus unserem Haus?

Dann melden Sie sich gerne zu einem oder mehreren themenspezifischen Newslettern an:



[newsletter-anmeldung-wp.fir.de](https://www.fir.de/newsletter-anmeldung-wp.fir.de)



1 Management-Summary

Geschäftskritischer Erfolgsfaktor oder notwendiges Übel: Stammdatenmanagement (SDM) ist eine Disziplin, die in der Praxis oft nicht den angemessenen Stellenwert erfährt. Meist wird eine schlechte Datenqualität zwar als Problem wahrgenommen, jedoch von den Unternehmen nicht behandelt. Das Resultat ist, dass Redundanzen und Inkonsistenzen in den Stammdaten weiter zunehmen und letztendlich auch die Prozesse, die auf die Daten zugreifen, ins Stocken geraten. Die mangelnde Konsequenz in der Datenpflege hat zwei Hauptursachen: Zum einen sind die konkreten Nutzenpotenziale mitunter nur schwer ersichtlich und der direkte Beitrag zum Unternehmenserfolg ist häufig nicht greifbar. Zum anderen mangelt es oftmals an geeigneten Lösungsansätzen oder deren konsequenter Umsetzung.

Ganzheitliches betriebliches Stammdatenmanagement erfordert sowohl strategische und organisatorische als auch informationstechnische Maßnahmen in Unternehmen. Das SDM muss mit der strategischen Planung des Unternehmens in Einklang gebracht werden. Geeignete Datenpflegeprozesse müssen definiert werden, die unter Einbindung von IT- und Fachabteilungen sowohl Verantwortlichkeiten als auch das Vorgehen beschreiben. Eine Stammdatenarchitektur muss gewählt werden, die aktuelle und korrekte Stammdaten innerhalb der IT-Landschaft gewährleistet. Geeignete Datenstrukturen müssen geschaffen werden, die sowohl die Anforderungen der Fachabteilungen abdecken als auch den benötigten Grad an Interoperabilität ermöglichen.

Die erste Version des Whitepapers Stammdatenmanagement wurde circa 1000 Mal heruntergeladen. Die vorliegende zweite Auflage wurde in vielen Bereichen aktualisiert und um einige neue Themenbereiche ergänzt.

Das Ziel des vorliegenden Whitepapers ist es, dem Leser einen Einblick in das Themengebiet Stammdatenmanagement zu gewähren. Anfangs werden hierfür das Stammdatenthema vorgestellt und die Grundlagen beschrieben. Ferner werden die einzelnen Handlungsfelder und Gestaltungsbereiche aufgezeigt. Ein weiterer Abschnitt widmet sich der Schnittstelle zwischen Big Data und Stammdaten. Letztendlich liefert das Whitepaper eine Vorgehensweise, um systematisch die Stammdatenqualität zu optimieren und nachhaltiges SDM im Unternehmen zu etablieren.



2 Motivation

Die Informationstechnologie (IT) ist unlängst zum kritischen Erfolgsfaktor für die Wettbewerbsfähigkeit von Unternehmen avanciert. Unternehmen aller Größenordnungen verlassen sich auf immer umfangreichere und komplexer werdende IT-Systeme zur Unterstützung ihrer Geschäftsprozesse. Systeme wie „Enterprise-Resource-Planning“ (ERP), Customer-Relationship-Management (CRM) und „Business-Intelligence“ (BI) sind aus dem Arbeitsalltag nicht mehr wegzudenken. Zudem bieten neue Technologien wie RFID, Sensorsysteme oder Smart Objects gerade im produzierenden Sektor weiteres Potenzial, werden jedoch auch zu einem stetigen Datenwachstum beitragen. Dieses Wachstum bedeutet für viele Unternehmen auch die Erschließung neuen Potenzials oder gar neuer Geschäftsmodelle.

Trotz der Sensibilisierung für das Thema Big Data wird das Thema der Datenpflege oftmals vernachlässigt. Studien zufolge haben nur ca. 40 Prozent der Unternehmen (klar) definierte Strukturen und Prozesse im SDM. Zwar hat ein Großteil der Unternehmen schon einmal Stammdatenprojekte durchgeführt, jedoch keine kontinuierlichen

Maßnahmen etabliert. Insbesondere die mangelnde Verfügbarkeit interner Ressourcen wird hierfür als Grund genannt (s. SCHEIBMAYER u. KNAPP 2014, S. 31). Eine Hauptursache dafür ist, dass eine schlechte Datenqualität erfahrungsgemäß nicht unmittelbar zum Stillstand in einem Unternehmen führt. Vielmehr handelt es sich bei der Verschlechterung der Datenqualität um einen schleichenden Prozess, der die reibungslosen Abläufe im Unternehmen, beispielsweise durch vermehrte Abstimmungen und Rückfragen, immer weiter verlangsamt. Die Analogie zur Redensart „Sand im Getriebe haben“ liegt insofern nahe. Ohne konsistentes und durchgreifendes Stammdatenmanagement gerät über die Zeit immer mehr „Sand in das Getriebe“, bis es durch zu viel Reibung letztendlich zum kompletten Stillstand kommt. So weit muss es jedoch gar nicht kommen. Es gibt viele Möglichkeiten, mangelhafter Stammdatenqualität mithilfe eines sinnvoll integrierten Stammdatenmanagements entgegenzuwirken. Eine der ersten Fragen, die hierfür beantwortet werden muss, wird im folgenden Abschnitt behandelt: Was sind eigentlich Stammdaten und was ist Datenqualität?



2.1 Stammdaten und Datenqualität

Stammdaten beschreiben die wesentlichen betrieblichen Objekte innerhalb eines Unternehmens wie Kunden, Produkte, Kunden und Lieferanten (s. Bild 1). Sie sind die Referenz für sämtliche anderen Daten im Unternehmen, sind aber selber unabhängig und haben üblicherweise eine geringe Änderungsfrequenz – im Gegensatz zu Bestandsdaten, welche beispielsweise die Mengen eines Objekts zu einem Zeitpunkt angeben und mit dem entsprechenden Stammdateneintrag (z. B. eine gewisse Anzahl einer bestimmten Schraube) verknüpft sind. Dasselbe gilt für die Bewegungs- und Änderungsdaten, welche Veränderungen der Stammdaten selbst oder deren Bestände auslösen. Folglich werden Stammdaten unternehmensweit in vielfältigen Geschäftsprozessen benötigt.

Stammdaten sind unterschiedlich strukturiert und unterliegen, je nach Einsatzgebiet, verschiedenen Anforderungen, z. B. hinsichtlich ihrer Zugänglichkeit. So sind Lieferantenstammdaten rein betriebsintern, Produkt- bzw. Artikeldaten können hingegen auch Kunden zugänglich gemacht werden.

Die Komplexität des Themas SDM wird insbesondere bei den Produktdaten deutlich, was sich leicht am Beispiel einer Schraube veranschaulichen lässt. Je nach Branche bestehen unterschiedliche Anforderungen an die zu beschreibenden Merkmale: Für Maschinen- und Anlagenbauer sind besonders technische Daten interessant, wie bspw. der Durchmesser, die Länge, das Material oder die Festigkeit. Unterliegt die Schraube Hochleistungsansprüchen, z. B. im Flugzeugbau, ist für den Kunden zusätzlich wichtig, dass sie zuvor auf Materialfehler überprüft und geröntgt wurde. In beiden Fällen muss die Stammdatenstruktur eine geeignete Beschreibung ermöglichen.

Generell ist die Qualität der Stammdaten das Maß, in dem die Daten die Realität beschreiben. Dies ist allerdings eine vereinfachte Darstellung, ähnlich der Feststellung „Die Qualität unserer Daten ist schlecht!“. Die Qualität kann in verschiedenen Bereichen des Unternehmens und für verschiedene Anwendungsfelder und Objekte anders definiert werden. Tabelle 1 zeigt mögliche Dimensionen der Stammdatenqualität. Insbesondere die Fehlerhaftigkeit, Unvollständigkeit und mangelnde Aktualität bereiten den Unternehmen dabei Probleme (S. SCHEIBMAYER U. KNAPP 2014, S. 19).

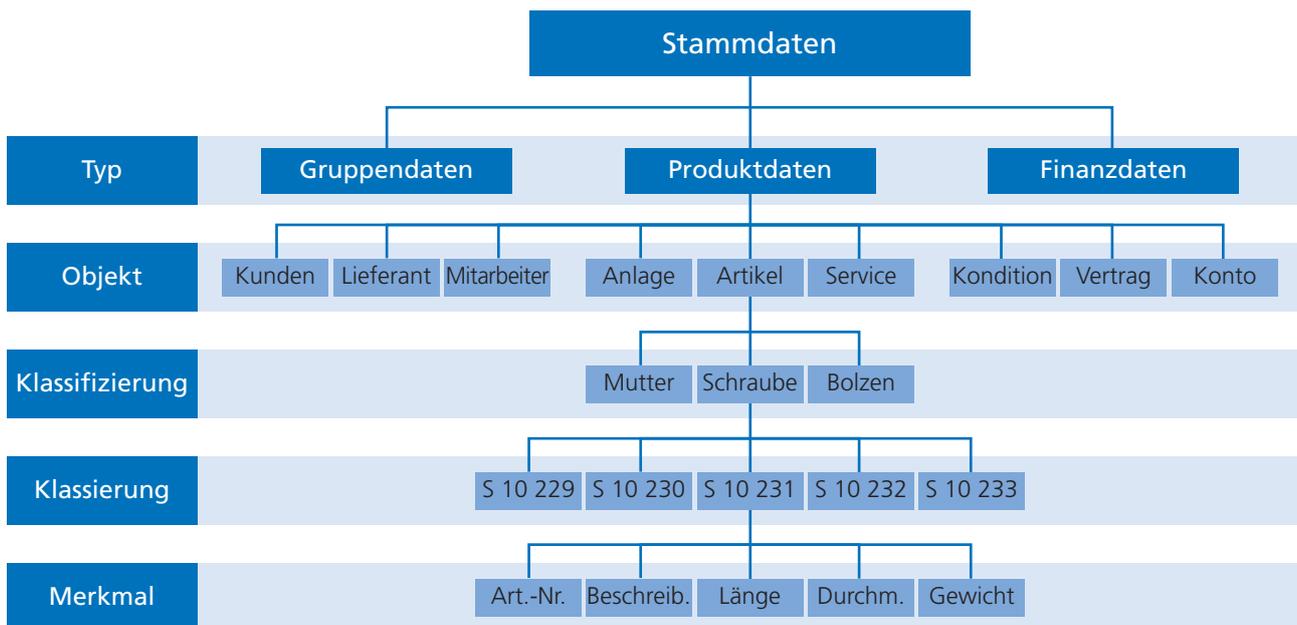


Bild 1: Stammdaten in Unternehmen (eigene Darstellung)

Dimensionen der Stammdatenqualität

- **Glaubhaftigkeit ...**
... (engl. *believability*) beschreibt das Maß, in dem Stammdaten als wahr und zuverlässig angesehen werden. Allgemeine Indikatoren für hohe Glaubhaftigkeit sind Zertifikate, die einen hohen Qualitätsstandard ausweisen, sowie der Aufwand, der für die Informationsgewinnung, -verbreitung und -überprüfung betrieben wurde.
- **Verfügbarkeit ...**
... (engl. *availability*) beschreibt, inwiefern Stammdaten in einer geforderten Zeitspanne zur Verfügung stehen. Ein Maß für die zeitliche Verfügbarkeit ist die Wartezeit zwischen dem Zeitpunkt der Informationsanforderung und der Bereitstellung durch die Quelle.
- **Aktualität ...**
... (engl. *currency* oder *timeliness*) beschreibt den Grad, zu dem Stammdaten zu einem bestimmten Zeitpunkt der Realität entsprechen. Dabei hat sich für Stammdaten eine Einordnung nach deren Alter als praktikabel erwiesen.
- **Fehlerfreiheit ...**
... (engl. *correctness*) beschreibt, inwiefern Stammdaten korrekt und frei von Fehlern sind. Insbesondere kann zwischen syntaktischer und semantischer Fehlerfreiheit unterschieden werden.
- **Angemessener Umfang ...**
... (engl. *accuracy*) beschreibt den Grad, in dem die Abbildung realer Objekte in den Stammdaten den Anforderungen der Nutzer genügt.
- **Vollständigkeit ...**
... (engl. *completeness*) beschreibt, inwiefern die dafür vorgesehenen Merkmale in einem Stammdatensatz beschrieben bzw. mit Ausprägungen belegt sind.
- **Strukturiertheit ...**
... (engl. *structuredness*) beschreibt den Grad, in dem die Stammdaten durch ein einheitliches Muster zur Beschreibung ihrer unabhängigen Objekte beschrieben werden.
- **Zugänglichkeit ...**
... (engl. *accessibility*) beschreibt, inwiefern Nutzer auf einfachem Weg auf die Stammdaten in den Anwendungssystem zugreifen und diese abrufen können.
- **Eindeutigkeit bzw. Redundanzfreiheit ...**
... (engl. *redundance-free* oder *consistency*) beschreibt das Maß, zu dem der Stammdatenbestand frei von Redundanzen (Dubletten) ist.
- **Verständlichkeit ...**
... (engl. *uniqueness* oder *interpretability*)
beschreibt den Grad, zu dem die Stammdaten durch alle Empfänger gleich verstanden werden.
- **Übersichtlichkeit ...**
... (engl. *clearness*) beschreibt, inwiefern Informationen in einem für den Nutzer passenden und leicht fassbaren Format dargestellt sind.
- **Konsistenz ...**
... (engl. *consistency*) beschreibt die Widerspruchsfreiheit der Stammdatenarten, d. h., ob mehrere für sich betrachtete Stammdatenobjekte auch im Zusammenhang korrekt sind.

Tabelle 1: Übersicht verschiedener Qualitätsdimensionen bei Stammdaten
(Quelle: SCHEIBMAYER U. KNAPP 2014, S. 18)

2.2 Auswirkung und Nutzenaspekte von Datenqualität

Eine mangelnde Datenqualität in den Stammdatenbeständen führt zu Leistungseinbußen der IT-Systeme und letztendlich dazu, dass die eigentlich zu unterstützenden Geschäftsprozesse ins Stocken geraten. Die Potenziale, die im Bereich Datenqualität brachliegen, sind enorm: ca. 40 Prozent der Unternehmen bewerten beispielsweise ihre Produkt- und Anlagendaten als unterdurchschnittlich, wobei kleine und mittlere Unternehmen (KMU) die Qualität durchschnittlich noch etwas besser einschätzen als große Unternehmen (s. SCHEIBMAYER u. KNAPP 2014, S. 16). Die Konsequenzen sind weitreichend; so werden u. a. 40 Prozent aller Bestellungen aufgrund von Stammdatenproblemen verzögert (vgl. HAUG ET AL. 2013, S. 237). Folglich kann auch die Integration von Big-Data-Anwendungen zu Problemen bzw. fehlerhaften Ergebnissen kommen, sofern die Datenbasis nicht ordentlich gepflegt ist.

Die ganzheitliche Anwendung von betrieblichem SDM erhöht die Qualität der Stammdaten und stellt sie nachhaltig sicher. Hierdurch kann hohes Nutzenpotenzial in unterschiedlichen Unternehmensbereichen entlang der Wertschöpfungskette erschlossen werden (s. Bild 2).

Planung

Präzises und korrektes Wissen über Produkte, Kunden und Finanzen eines Unternehmens sind Grundvoraussetzungen für richtige Geschäftsentscheidungen. Hohe Datenqualität erweist sich an dieser Stelle als Grundlage für die strategische Geschäftsführung. Das Risiko von Fehlentscheidungen aufgrund von falschen Daten wird stark reduziert.

Einkauf und Beschaffung

Studien zufolge können 80 Prozent aller Unternehmen das noch ungenutzte Potenzial im Bereich Einkauf und Beschaffung aufgrund unzureichender

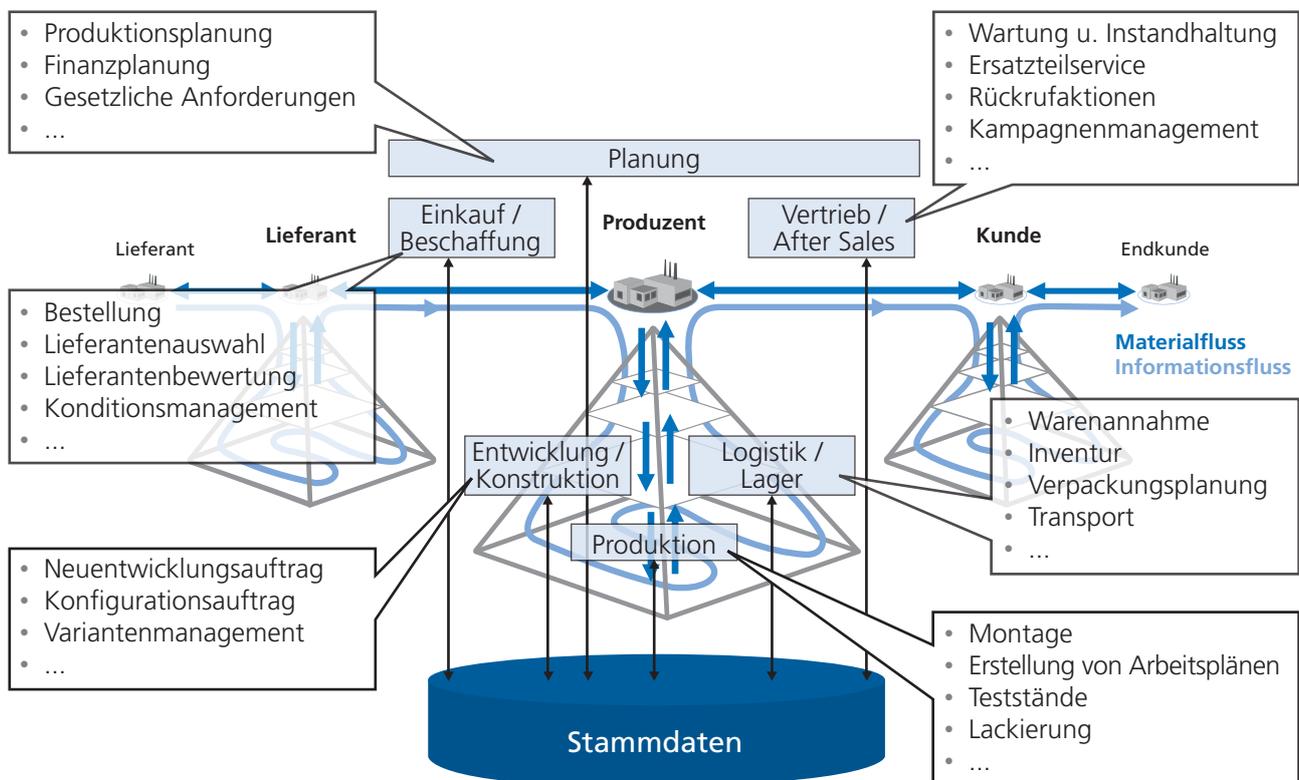


Bild 2: Auswirkungen und Nutzenpotenziale von SDM nach Unternehmensbereichen (eigene Darstellung)

Daten- und Informationsqualität nicht ausschöpfen (s. RUMP 2009, S. 8f). Dieser Umstand ist meist auf niedrige Datenqualität zurückzuführen. Eine einheitliche Sicht auf die Materialdaten ermöglicht die Bündelung von Einkaufsaktivitäten, während die konsistente Darstellung der Lieferantendaten die Lieferantenkonsolidierungen unterstützt. Zusätzlich können Lagerbestände durch die Vermeidung von Doppelbeständen optimiert werden.

Produktion

Stimmen z. B. Mengenangaben in Stücklisten nicht, werden zu viele oder zu wenige Vorprodukte gefertigt. Sauber gepflegte Stammdaten reduzieren das Risiko von Fehlproduktionen deutlich. Darüber hinaus unterstützt ein funktionierendes SDM die Einhaltung von Compliance-Anforderungen, beispielsweise hinsichtlich Chargenverfolgungen. Insgesamt wird die Produktionsplanung durch konsistente Daten im ERP verbessert.

Logistik und Lager

Eine klassische Auswirkung von schlechter Datenqualität ist die doppelte Lagerhaltung, welche durch (leicht) unterschiedliche Produktbezeichnungen verursacht wird. Folglich hat das Unternehmen eine unnötig erhöhte Kapitalbindung. Die Auswirkungen können sich bis auf die Produktion auswirken, falls ein Zulieferprodukt neu beschafft werden muss, obwohl es sich unter anderer Bezeichnung noch im Lager befindet. Insgesamt bilden gepflegte Stammdaten eine solide Basis für Automatisierungsvorgänge, wie beispielsweise die Kommissionierung in der Logistik und in Lagern.

Vertrieb und After-Sales

Erhöhte Datenqualität beugt Fehlern bei Bestellungen und Lieferungen vor und senkt dadurch die Anzahl der Kundenreklamationen und Retouren. Dies führt zu geringeren Fehlerbehebungskosten, einer Entlastung des Kundendienstes und letztendlich zu höherer Kundenzufriedenheit. Weiteres Potenzial kann durch die einfachere Verwendung von (elektronischen) Produktkatalogen erschlossen werden.

2.3 Herausforderungen beim Stammdatenmanagement

Die Umsetzung von effizientem betrieblichem Stammdatenmanagement stellt Unternehmen jeder Größenordnung vor Herausforderungen. Die Unterstützung des Managements sowie die Beteiligung der IT-Abteilung und der Fachbereiche sind dabei unerlässlich für den Erfolg der Datenpflege. SDM muss in der IT-Strategie des Unternehmens verankert werden, um so ein Bewusstsein für die Relevanz dieses Themas zu schaffen. Ferner muss sichergestellt werden, dass das Fachwissen der verschiedenen Abteilungen – auch werksübergreifend – durch die Stammdatenstruktur abgebildet und ein vorher definierter Qualitätsstandard eingehalten wird. Zu diesem Zwecke sind vor allem organisatorische Strukturen – wie Zuständigkeiten der Datenpflege und entsprechende Pflegeprozesse – zu definieren.

Das folgende Kapitel gibt eine Übersicht über die Gestaltungsfelder von SDM in Unternehmen und beschreibt Methoden, um diese optimal zu gestalten.

3 Unternehmensweites Stammdatenmanagement

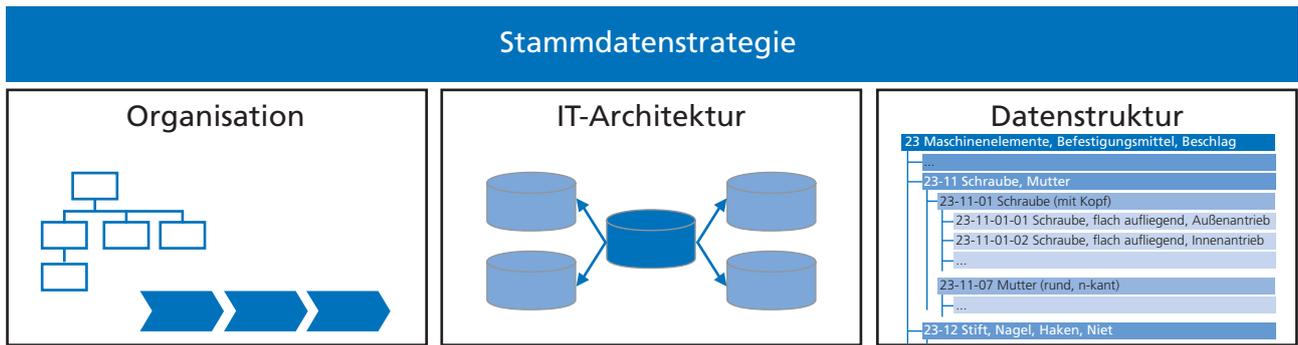


Bild 3: Handlungsfelder des unternehmensweiten SDMs (eigene Darstellung)

Entgegen vielen Meinungen ist Stammdatenmanagement kein reines IT-Thema. Eine nachhaltige Umsetzung verlangt nach einer ganzheitlichen Lösung, die von der jeweiligen Geschäftsführung, der IT-Abteilung und den Fachbereichen entwickelt und umgesetzt werden muss. Es zeigt sich, dass insbesondere die Unterstützung der Geschäftsführung einen nicht zu vernachlässigenden Erfolgsfaktor darstellt (s. SCHEIBMAYER U. KNAPP 2014, S. 44). In Bild 3 werden die verschiedenen Gestaltungsfelder des unternehmensweiten Stammdatenmanagements veranschaulicht.

Die folgenden Kapitel geben einen kurzen Überblick über Aufgaben innerhalb der einzelnen Handlungsfelder.

3.1 Stammdatenstrategie

Die strategische Planung soll allgemein die Realisierung der Unternehmensziele unterstützen. In diesem Kontext ist auch das Thema Stammdatenmanagement zu sehen. Es hat vielfältige Auswirkungen auf Geschäftstreiber, wie Geschäftsprozesse, Qualitäts- und Risikomanagement, Kundenzufriedenheit etc. und ist daher unter strategischen Aspekten zu betrachten (vgl. OTTO U. HÜNER 2009, S. 19ff). Aus diesem Grund muss jede Stammdatenmanagementinitiative mit der allgemeinen strategischen Planung synchronisiert und der Beitrag zur Erreichung der Unternehmensziele verdeutlicht werden. Dies hebt auch das Thema Stammdatenmanagement aus einer reinen IT-Verantwortung auf eine unternehmensweite Perspektive.

In diesem Zusammenhang fällt auch oftmals der Begriff der Data-Governance oder des Data-Governance-Konzepts. Eine einheitliche Definition für diesen Begriff existiert in der Literatur nicht. Ausgehend von der Bedeutung der Wörter heißt es soviel wie „Macht über Daten“. Fest steht auch, dass eine Data-Governance unternehmensweite Leitplanken für die effektive Benutzung von Daten vorgibt und somit das Erreichen der strategisch definierten Unternehmensziele unterstützt. Dies betrifft also neben dem reinen Managen von Daten auch Prozesse, Mitarbeiter, die verfügbaren Technologien und Strukturen im Unternehmen. Somit sollte auch sichergestellt sein, dass die entsprechenden Verantwortlichen in die relevanten Themen involviert sind und somit eine effiziente Nutzung der Daten möglich ist.

3.2 Organisatorische Verankerung

Eine wichtige Charakteristik des Stammdatenmanagements ist der Prozessbezug; um nachhaltig eine gute Stammdatenqualität zu erreichen, müssen alle operativen Prozesse eines Unternehmens definiert sein, die das Anlegen, Ändern oder Deaktivieren von Stammdaten betreffen. Um ein umfassendes Organisationskonzept entwickeln zu können, müssen die in Bild 4 aufgezeigten Schritte durchgeführt werden.

Als erster Schritt müssen Rollen definiert werden, die für das Stammdatenmanagement im Unternehmen relevant sind. Klassische Beispiele für mögliche Rollen sind: Datenowner, Auftraggeber/Anforderer, Prüfstelle, durchführende Stelle oder Kommunikationsstelle.

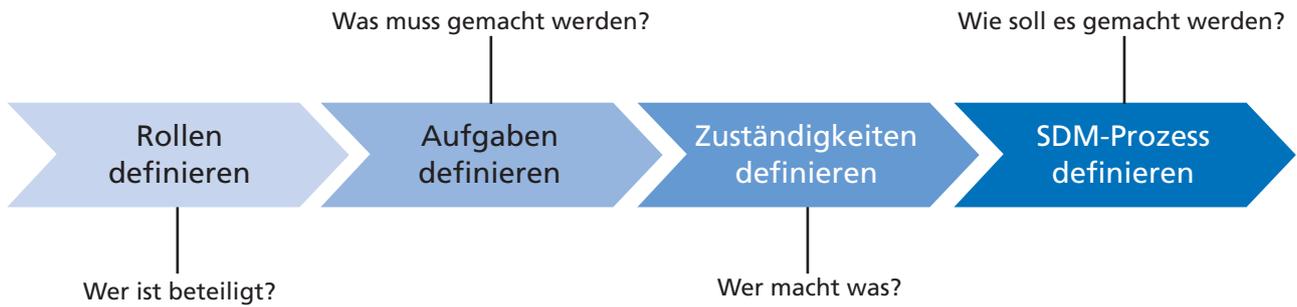


Bild 4: Der Weg zur Stammdatenorganisation (eigene Darstellung)

In einem nächsten Schritt müssen Aufgaben definiert werden, die für ein funktionierendes SDM erfüllt werden müssen. Diese orientieren sich zum einen am Lebenszyklus von Stammdaten, der sich beschreiben lässt durch deren Anlage, Änderung, Deaktivierung und die Archivierung bzw. Löschung. Zum anderen beziehen sich diese jedoch auch auf unterstützende Aufgaben, wie die Analyse und Bereinigung von Stammdaten.

Sind die Aufgaben hinreichend definiert, müssen sie den im ersten Schritt beschriebenen Rollen im Unternehmen zugeordnet werden. Darüber hinaus ist zu klären, in welchem Umfang welche

Rolle an den Aufgaben beteiligt ist. Für die Visualisierung dieser Zuordnung ist die Erstellung einer Zuständigkeitsmatrix ein adäquates Mittel (s. Bild 5). Sie zeigt deutlich, bei welcher Rolle die Verantwortung liegt, die Aufgabe durchzuführen, und welche Rollen nur mitwirken oder informiert werden müssen.

Auf einer anderen Ebene kann durch Zuständigkeitsmatrizen z. B. auch die Verantwortung einzelner Stammdatentypen zu Organisationseinheiten geklärt werden. So könnten beispielsweise alle Zuständigkeiten bezüglich Stammdaten hinsichtlich Fertigwaren beim Vertrieb liegen und der Einkauf verantwortlich für Daten zu Rohstoffen sein.

		Rolle					
		Besteller/ Auftraggeber	Prüfstelle	Datenowner- Management	Datenowner- Prozess	Mutations- stelle	Kommuni- kationsstelle
Aufgabe	Auslösung einer Veränderung	R					
	Überprüfung der Richtigkeit		R				
	Freigabe			R, A	C		
	Veränderung der Daten					R	
	Unternehmens- weite Publikation	I	I	I	I	I	R

Legende: R = verantwortlich, A = haftbar, C = Rücksprache, I = informiert

Bild 5: Zuständigkeitsmatrix (i. A. a. KHAN, 2014)

Schließlich sollten auch die Prozesse, die hinter den Aufgaben stehen, dokumentiert und somit im Unternehmen transparent gemacht werden. Im Idealfall lassen sich diese Prozessdokumentationen in ablauffähige Workflows transformieren. Bild 6 zeigt beispielhaft, wie der Workflow einer Materialstamm-Anlage aussehen könnte. Ein Konstrukteur würde in diesem Fall eine Anfrage zur Anlage eines neuen Stammdatensatzes ausfüllen und elektronisch an eine übergeordnete Prüfungsstelle schicken. Diese prüft zunächst, ob alle formalen Kriterien hinsichtlich Vollständigkeit der Angaben und Eindeutigkeit der Bezeichnungen erfüllt sind. Darüber hinaus wird eine fachlich-inhaltliche Prüfung durchgeführt, die sicherstellen soll, dass die Neuanlage notwendig und sinnvoll ist. Besteht z. B. die Möglichkeit, einen ähnlichen Teil mit gleichen Eigenschaften

zu nutzen, würde die Neuanlage verweigert. So können kostspielige Neuentwicklungen von Teilen vermieden und ggf. Mengenpotenziale beim Einkauf erschlossen werden. Ist der Antrag auch inhaltlich geprüft und freigegeben, muss er umgesetzt und der Antragsteller informiert werden.

Das Beispiel macht auch deutlich, dass es in der organisatorischen Ausgestaltung meist verschiedenen Alternativen gibt: So kann die formale Prüfung durch eine organisatorische Stelle im Unternehmen, d. h. von einer Person oder durch ein IT-System, durchgeführt werden, z. B., indem der Antragsteller beim Ausfüllen einer Maske bereits auf Unzulänglichkeiten hingewiesen wird. Wichtig ist es, deutlich zu machen, dass es nicht darum geht, Mitarbeiter zu kontrollieren, sondern Fehler möglichst früh zu vermeiden.

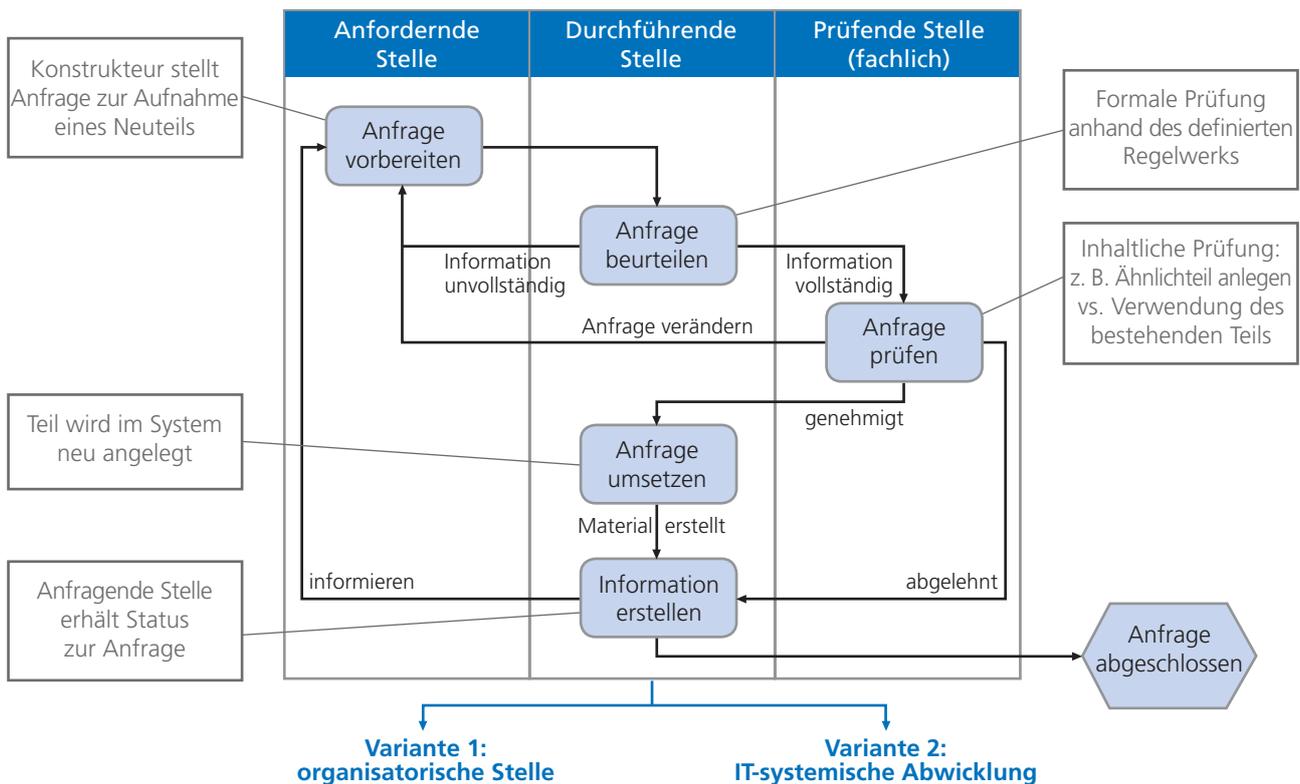


Bild 6: Beispielprozess/-workflow (eigene Darstellung)

3.3 IT-Architekturen für das Stammdatenmanagement

Für die unternehmensindividuelle IT-Architektur des SDMs gibt es verschiedene Modelle, abhängig von den jeweiligen Anforderungen. Die Anforderungen an die Lokalität bzw. Zentralität der Daten und ihr Harmonisierungsgrad sind dabei entscheidend für die Auswahl. Die Architektur prägt die zukünftige Arbeit mit den Stammdaten bezüglich der Komplexität der Datenhaltung, Datenverteilung, des organisatorischen Aufwands, der Konsistenz, Redundanz und Aktualität der Datenbestände. Die passende Verwaltungsstruktur für ein Unternehmen muss daher in der Planungsphase ermittelt werden. Grundlage hierfür sind die Anforderungen der beteiligten Anwendungen und das Zusammenspiel von IT-Architektur und Datenstruktur. Allgemein lassen sich

typische Ausprägungen der IT-Architektur für das SDM hinsichtlich des Zentralisierungs- und Harmonisierungsgrades unterscheiden.

Werden Stammdaten zentral verwaltet, sind Organe zur Überwachung der Datenqualität einfacher zu installieren. Der gesamte Datenbestand ist leichter zu harmonisieren, die Konsistenz ist gewährleistet. Aufgrund der notwendigen Verteilung der Daten an die Anwendungen entstehen jedoch Verzögerungen, wodurch die Aktualität der Informationen beeinträchtigt wird.

Dezentrale Strukturen haben den Vorteil hoher Aktualität durch lokale Datenverwaltung, bergen aber die Gefahr der Inkonsistenz und Redundanz. Dadurch ist insbesondere der Verwaltungsaufwand höher, da alle Anwendungen eigenständig ihre Datenqualität kontrollieren müssen.

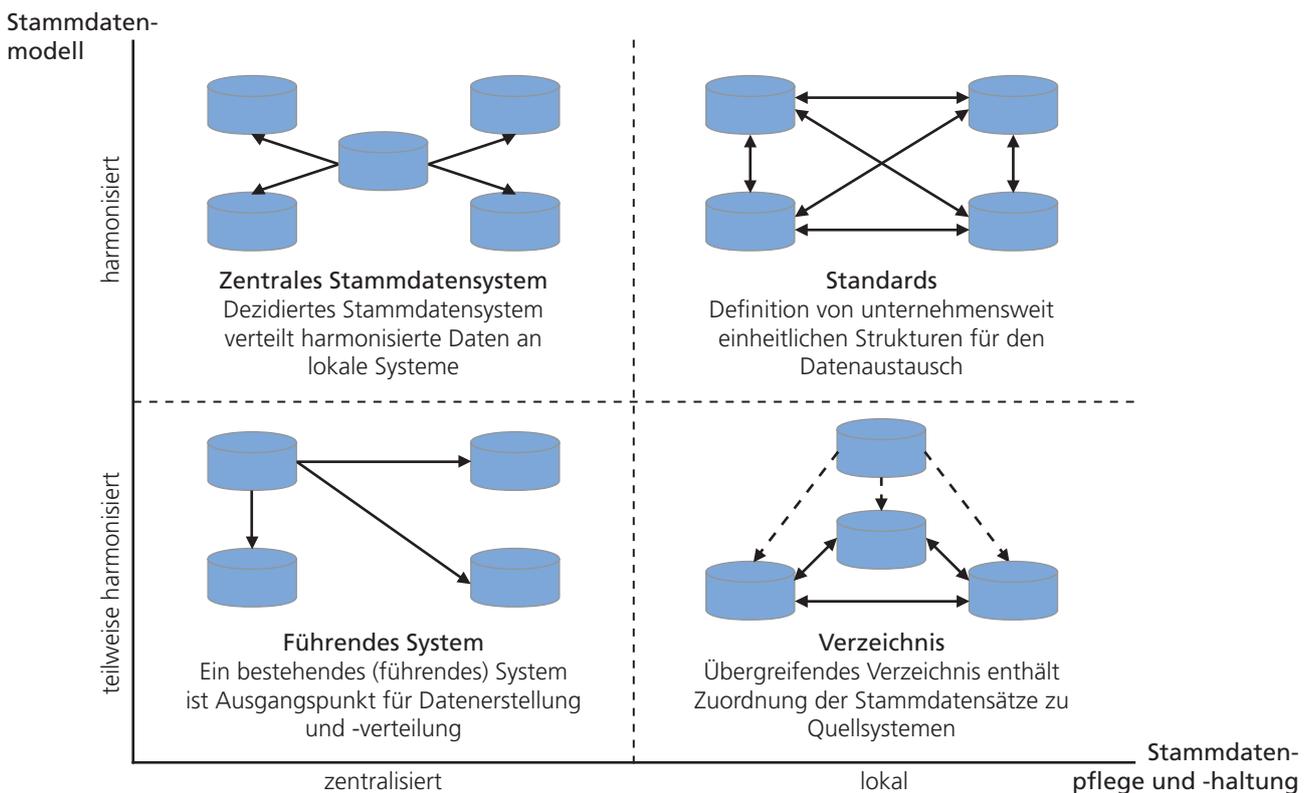


Bild 7: Übersicht der Architekturen (eigene Darstellung)

Der Harmonisierungsgrad der Stammdaten entscheidet über die Interoperabilität verschiedener Systeme. Die vollständige Harmonisierung des gesamten Datenbestands ist aufwendig und erfordert gute Planung und eine umfassende Analyse aller Anforderungen. Neben einer einmaligen Datenstrukturierung und ggf. Datenbereinigung muss die nachhaltige Datenpflege für neue Daten und Datenänderungen gewährleistet werden. Ein geringer Harmonisierungsgrad hat zur Folge, dass die Stammdaten für den Austausch zwischen verschiedenen Systemen über Ergänzungen, Converter, Mapping von Schlüsseln und Attributen aufbereitet werden müssen. Verändern sich die Anforderungen und Datenflussstrukturen, müssen diese Hilfsmittel angepasst bzw. neu erstellt werden.

Bild 7 zeigt die Architekturen abhängig vom Harmonisierungsgrad und der Zentralisierung. Die einzelnen Knoten in der Abbildung sind Anwendungen. Im Folgenden werden die vier wesentlichen Architekturen vorgestellt (vgl. LOSER 2008 et al., S.2 ff.).

Führendes System

Ein führendes System entsteht, wenn eine bestehende Anwendung mit der Datenpflege betraut wird. Das führende System übernimmt dabei die Datenverteilung. Die Daten werden redundant in den einzelnen Anwendungen gespeichert und gegebenenfalls mit lokalen Attributen ergänzt. Es gibt keinen einheitlichen Identifikationsschlüssel für die Stammdaten, daher müssen die Identifikationsschlüssel der Anwendungen aufeinander abgestimmt werden (Mapping). Der Vorteil dieser Architektur liegt in der Unabhängigkeit der Anwendungen, allerdings ist die Erweiterung um zusätzliche Anwendungen aufwendig, da ein entsprechendes Mapping eingerichtet werden muss.

Zentrales Stammdatensystem

Bei einem zentralen Stammdatensystem wird, im Gegensatz zum führenden System, eine zentrale und separate Anwendung mit der Verteilung und Pflege der Stammdaten betraut. Die Stammdaten sind harmonisiert und einheitlich und werden ausschließlich vom zentralen System an die Anwendungen verteilt. Für alle Daten gibt es einen einheitlichen Identifikationsschlüssel, Mappings

sind nicht notwendig. Die Anbindung weiterer Anwendungen ist einfach, allerdings führt die Push-Verteilung dazu, dass bei der Aktualisierung von Daten Verzögerungen entstehen.

Verzeichnis

In einem Verzeichnis wird der Speicherort anwendungsübergreifender Daten dokumentiert. Wenn eine Anwendung Daten benötigt, wird eine Anfrage an das Verzeichnis gestellt, die Antwort des Verzeichnisses benennt den Speicherort der Daten. Die anfragende Anwendung ruft nun die Daten von der entsprechenden Anwendung ab und ist für die Migration der Daten verantwortlich. Die Datenspeicherung erfolgt dezentral, es gibt keine festgelegten Datenverteilungswege. Die autonome Datenverwaltung der Anwendungen hat den Nachteil, dass es keinen einheitlichen Prozess der Datenerstellung und Datenpflege gibt. Bei der Anforderung von Daten entsteht Migrationsaufwand.

Die Identifikationsschlüssel der einzelnen Anwendungen werden im Verzeichnis in einen globalen Schlüssel untergeordnet. Da Daten nicht über ein zentrales System aktualisiert, sondern Veränderungen direkt in den Anwendungen durchgeführt werden, sind die angefragten Daten immer aktuell. Es entstehen keine Verzögerungen durch Neuverteilung veränderter Daten.

Standards

Die Einführung von unternehmensweiten Standards befähigt alle Anwendungen zur Kommunikation miteinander. Eine Integrationsschicht stellt sicher, dass alle Anwendungen mit den Stammdaten gemäß dem verwendeten Standard verfahren. Der Standard definiert dabei eine Struktur für die Stammdaten sowie globale Attribute. Die Daten werden dezentral in den Anwendungen gespeichert, standardisiert geführt und können daher ohne Migrationsaufwand zwischen den Anwendungen direkt angefragt und ausgetauscht werden.

Es gibt keinen Aufwand für Mapping und Migration, aber auch keine zentrale Erfassung und Kontrolle der Daten und keine vorgeschriebenen Datenverteilungswege, weshalb Redundanzen möglich sind. Da die Daten nicht zentral verwaltet werden, gibt es auch keinen einheitlichen globalen Identifikationsschlüssel der Daten.

	Vorteile	Nachteile
Führendes System	<ul style="list-style-type: none"> ⊕ größtenteils harmonisierter Datenbestand bei gleichzeitig unabhängigen Anwendungen ⊕ Konsistenz des Datenbestands gewährleistet 	<ul style="list-style-type: none"> ⊖ Datenergänzung und Mappings von Keys und Attributen erforderlich ⊖ redundante Datenhaltung ⊖ Verzögerungen bei Datenverteilung möglich
Zentrales Stammdatenmanagement	<ul style="list-style-type: none"> ⊕ harmonisierter Datenbestand ⊕ Redundanzfreiheit und Konsistenz des Datenbestands gewährleistet ⊕ geringer Organisationsaufwand der Datenpflege, häufig workflowgestützt 	<ul style="list-style-type: none"> ⊖ Verzögerungen bei Datenverteilung möglich, dadurch geringere Aktualität der Daten ⊖ hoher initialer Aufwand zur Schaffung geeigneter Strukturen
Verzeichnis	<ul style="list-style-type: none"> ⊕ keine Datenverteilung notwendig ⊕ hohe Aktualität des Datenbestands ⊕ geringer Umsetzungsaufwand 	<ul style="list-style-type: none"> ⊖ kein harmonisiertes Datenmodell ⊖ Redundanzen und Inkonsistenzen durch dezentrale Datenhaltung ⊖ kein „Single Point of Truth“ ⊖ kontinuierlich hoher Pflegeaufwand
Standards	<ul style="list-style-type: none"> ⊕ harmonisiertes Datenmodell ⊕ keine Datenverteilung notwendig ⊕ hohe Aktualität des Datenbestands 	<ul style="list-style-type: none"> ⊖ Inkonsistenzen durch dezentrale Datenhaltung ⊖ anfällig für Redundanzen ⊖ kein „Single Point of Truth“ ⊖ kontinuierlich hoher Pflegeaufwand

Bild 8: Vor- und Nachteile der verschiedenen SDM-Architekturen (eigene Darstellung)

3.4 Stammdatenstruktur

Ein hoher Harmonisierungsgrad und die damit verbundene Interoperabilität der Anwendungen erfordert eine homogene Stammdatenstruktur. Zu diesem Zweck werden Stammdatenklassifikationen verwendet, die einheitliche Merkmalschemata verwenden und darauf aufbauend eine Klassierung der Stammdatenobjekte ermöglichen.

Geeignete Merkmalschemata für die zukünftige Stammdatenklassifikation können durch die Analyse der vorhandenen Daten nach Unterscheidungsmerkmalen und kennzeichnenden Attributen erarbeitet werden. So tauchen beispielsweise gerade im Bereich der Artikel Merkmale wie ‚Länge‘, ‚Gewicht‘ oder ‚Dichte‘ auf. Diese Analyse sollte sowohl das Detailwissen über die Stammdatenobjekte als auch das Prozesswissen berücksichtigen. Eine reine Top-down-Analyse vernachlässigt meist Details (z. B. von Produkten), während eine reine Bottom-up-Analyse übergreifende Prozesse

vernachlässigt und die Kompatibilität zum Datenaustausch gefährdet. Die Mischung aus Top-down und Bottom-up hat sich in der Praxis als vorteilhaft erwiesen.

Auf Basis der identifizierten Merkmale können anschließend Stammdatenklassen abgeleitet werden. Ein Beispiel wäre eine Klasse ‚Schraube‘ mit den Merkmalen ‚Länge‘, ‚Durchmesser‘ und ‚Gewicht‘. Die Zusammensetzung der Merkmale sollte die Klasse von anderen Klassen klar getrennt sein. Sämtliche Stammdatenobjekte können anschließend klassiert, d. h. Klassen zugeordnet und über die Merkmale beschrieben werden. Sollten Restobjekte entstehen, muss eine adäquate Lösung diese integrieren oder es müssen zusätzliche Strukturen geschaffen werden.

Die finale Klassifikation kann anschließend durch einen geeigneten Klassifikationsstandard oder aber durch eine individuelle Datenstruktur realisiert werden. Die Wahl sollte abhängig von der

Unternehmenssituation und der verfolgten Strategie geschehen: So empfiehlt sich beispielsweise für einen Großhändler im Bereich Fast-Moving-Consumer-Goods (FMCG) der Einsatz eines Klassifikationsstandards zum schnellen Stammdatenaustausch, während ein Einzelfertiger aus dem Bereich Maschinen- und Anlagenbau aus Produktions- und Wettbewerbsgründen eher auf eine individuelle Struktur bauen könnte.

Klassifikationsstandards

Klassifikationsstandards werden vornehmlich verwendet, um den system- und vor allem unternehmensübergreifenden Stammdatenaustausch zu ermöglichen. Sie garantieren eine einheitliche Datenstruktur und dadurch die Interoperabilität zwischen unterschiedlichen Systemwelten. Bild 9 zeigt einige häufig verwendete Standards. Besondere Bedeutung kommt dem Standard eCl@ss zu, der international verwendet wird und durch die Einarbeitung branchenspezifischer Standards (bau:class, proficl@ss, PROLIST u.v.m.) immer größere Verbreitung findet.

Ausgehend von den ersten Ansätzen des elektronischen Datenaustauschs sind Klassifikationsstandards oft historisch gewachsen und werden stetig erweitert. Sie vereinen die beschreibenden Merkmale, die von Unternehmen an verschiedenen Stellen der Wertschöpfungskette benötigt werden. Daher sind sie oft kleinste gemeinsame Nenner dieser Anforderungen. Bei der Wahl des passenden Standards sollte daher berücksichtigt

werden, dass die spezifischen Anforderungen an die Stammdatenobjekte aus den Fachabteilungen abgedeckt werden. Werden diese nicht stark genug berücksichtigt, führt dies zu Fehlern und Problemen in den Prozessen. Viele Standards bieten aus diesem Grund eine Vielzahl an optionalen Feldern an, wodurch die Flexibilität erhöht wird. Bei deren Verwendung sollte jedoch darauf geachtet werden, dass der Datenaustausch weiterhin ohne Konvertierung möglich ist.

Individuelle Datenstrukturen

Individuelle Lösungen kommen zum Einsatz, wenn kein geeigneter Klassifikationsstandard existiert oder der Einsatz eines Klassifikationsstandards aus anderen Gründen nicht erwünscht ist. Sie bilden die Objekte innerhalb des Unternehmens meist sehr genau ab und berücksichtigen auch die spezielle Terminologie innerhalb des Unternehmens. Dadurch fördern sie ein unternehmensweites Verständnis über die Stammdatenobjekte und garantieren gleichzeitig die Individualität der Produkte. Individuelle Datenstrukturen besitzen jedoch auch eine Reihe von Nachteilen. Sie sind meist proprietär und erschweren die Kommunikation mit Kunden und Lieferanten. Für den automatisierten Datenaustausch müssen Converter geschaffen werden, die Daten von der Unternehmenssystematik in die Zielsystematik übersetzen. Die fehlenden Informationen zur Konvertierung müssen ergänzt werden; dies ist aber nur eingeschränkt automatisiert möglich, was einen steigenden Nachbearbeitungsaufwand zur Folge hat.

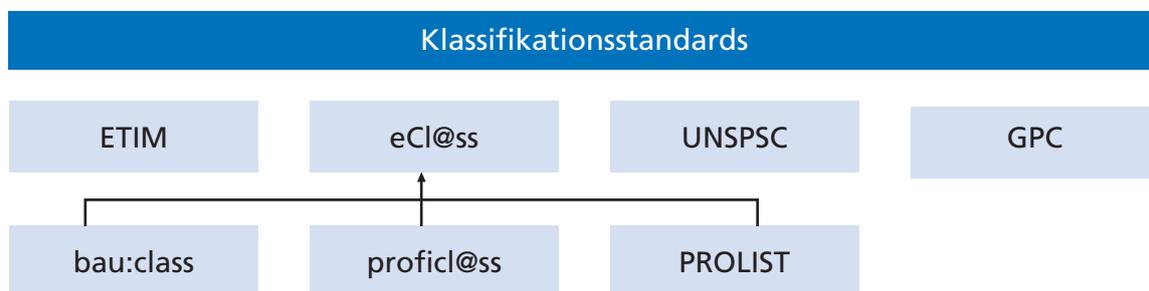


Bild 9: Übersicht gängiger Klassifikationsstandards (eigene Darstellung)

4 Vorgehen bei Stammdatenmanagementprojekten

Aufgrund der oben beschriebenen, komplexen Gestaltungsfelder stellt sich für Unternehmen die Frage, wie das unternehmensweite SDM nachhaltig verbessert werden kann. So individuell ein Stammdatenprojekt auch für ein jeweiliges Unternehmen aussehen kann, generell besteht ein typisches Stammdatenmanagement-Projekt aus vier aufeinander aufbauenden Phasen (s. Bild 10): *Datenanalyse, Datenstrukturierung, Datenharmonisierung* sowie *Organisation und Prozesse*.

Die erste Phase besteht aus der Analyse der bestehenden Datenlandschaft. Dabei werden die existierenden Datenstrukturen und -inhalte erfasst und gleichartige Objekte, Merkmale und Werte einander zugeordnet sowie syntaktisch und semantisch analysiert. Das Ziel ist eine Übersicht über die Qualität des Datenbestands mit Fokus auf Dubletten und Inkonsistenzen.

Im Rahmen der zweiten Phase (Datenstrukturierung) wird anschließend die Soll-Datenstruktur definiert, d. h. insbesondere eine Klassifizierung und Klassierung der Stammdaten vorgenommen

und Varianten abgeleitet. Hierbei kann auf standardisierte Klassifikationen zurückgegriffen werden, z. B. eCI@ss für Materialstammdaten. Die entstehende einheitliche Datenstruktur dient als Basis für das weitere Vorgehen.

In der dritten Phase (Datenharmonisierung) wird der Datenbestand bereinigt und auf die neue Struktur übertragen. Dabei werden insbesondere doppelte Datensätze vereinheitlicht bzw. eliminiert. Dieser Migrationsvorgang muss kontinuierlich kontrolliert werden, um eine hohe Datenqualität gewährleisten zu können. In der abschließenden vierten Phase werden nachhaltige Organisationsstrukturen für Veränderungen in den nun harmonisierten Datenbeständen geschaffen und in der Unternehmensstruktur verankert. Dies beinhaltet insbesondere die Bestimmung zuständiger Rollen, die Identifizierung notwendiger Aufgaben und letztendlich die Verteilung von Verantwortlichkeiten. Darauf aufbauend werden einheitliche Workflows definiert, die das Vorgehen bei Datenanlage, -modifikation und -deaktivierung eindeutig beschreiben und die Datenqualität nachhaltig sicherstellen.

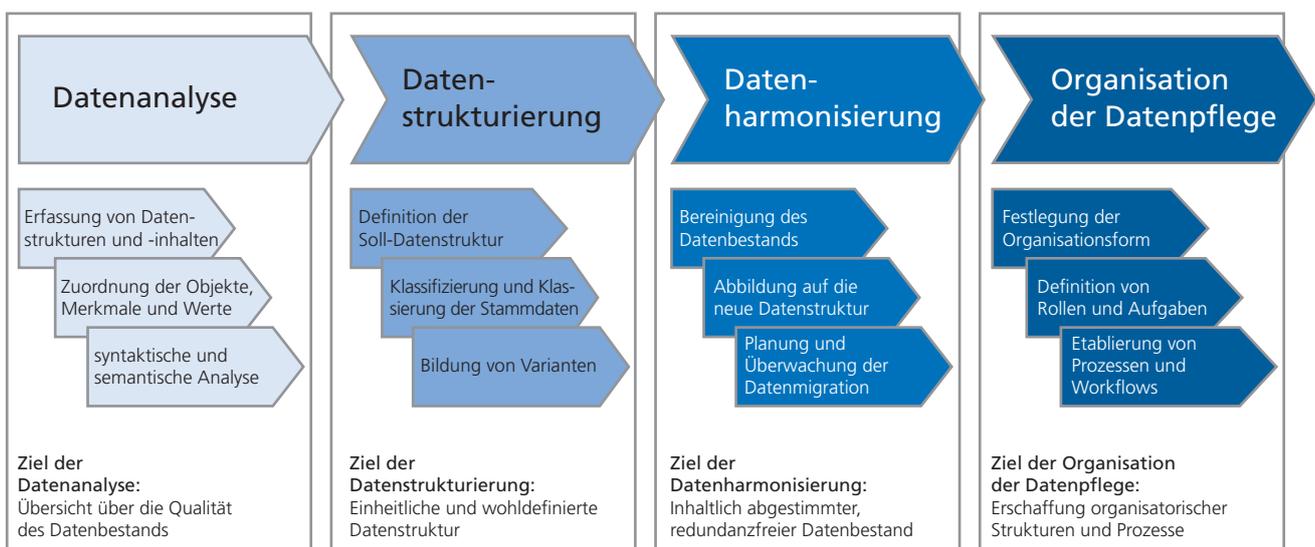


Bild 10: Die Phasen eines Stammdatenmanagement-Projekts (eigene Darstellung)

5 Stammdaten und Big Data / Smart Data

Den Begriff *Big Data* kennen wohl die meisten. Auch Smart Data ist noch vielen geläufig. Nahezu jeder kann sich unter diesen Schlagwörtern etwas vorstellen. Fragt man, was Big Data sei, werden oft die folgenden oder ihnen ähnliche Begriffe genannt, wie „Größe“, „Datenmassen“, „Echtzeitanalyse“ oder „Unstrukturiertheit“. Doch wann sind Datenmengen groß? Wie schnell müssen Daten analysiert werden, um von einer Echtzeitanalyse sprechen zu können? Wann sind Daten strukturiert oder unstrukturiert bzw. welcher Grad an Strukturiertheit ist für welche Analyse erforderlich?

Daher stellen sich uns die Fragen:

„Wie sind Big Data und Smart Data eigentlich definiert?“

„Wie genau unterscheiden sie sich voneinander?“

„Was sind ihre Vor- und Nachteile?“

Letztendlich sind diese Fragen allgemeingültig und endgültig nicht zu beantworten, allerdings soll im folgenden eine Kategorisierung vorstellen, wie man die Konzepte Big Data und Smart Data voneinander abgrenzen kann, um diesen Wörterdschungel ein wenig zu lichten. Dabei ist wichtig, dass eine Datenmenge nicht als „big“ oder „smart“ bezeichnet wird, sondern das Konzept der Datenverarbeitung von den Datenquellen bis hin zur daraus gewonnenen Information als Big-Data- oder Smart-Data-Konzept klassifiziert wird. Eine Datenmenge anhand ihrer Größe oder Komplexität als Big Data oder Smart Data zu bezeichnen, ist impraktikabel, da die Datenmengen kontinuierlich wachsen.

„Die Datenmengen und damit die Probleme steigen explosionsartig. Schon sprechen erste Experten nicht mehr von Big Data, sondern Huge Data“. Harald Weiss (s. WEISS 2015).

Beachtet man dabei noch, dass 90 Prozent aller weltweit gespeicherten Daten erst in den letzten zwei Jahren generiert wurden (s. WEISS 2015), so wird der Begriff „Huge Data“ in wenigen Jahren auch wieder zu klein werden und von beispielsweise „Humongous Data“ abgelöst werden. Statt also regelmäßig einen Begriff auszutauschen, ist es sinnvoller, ein Konzept unabhängig von der Datenmenge als Big-Data- oder Smart-Data-Konzept zu bezeichnen.

Genau wie die bei der Big-Data-Thematik oft implizit gemeinten Bewegungsdaten wachsen auch Stammdaten stetig in Größe und Komplexität an. Suchvorgänge werden immer aufwendiger, Strukturen immer komplexer. Zumeist liegen Stammdaten zwar in strukturierter Form vor, aber da es nicht einen zentralen Speicherort für jede Art von Datum gibt, sondern gleichartige Daten in verschiedenen Systemen gespeichert werden, liegt auch in diesen Daten eine Art Unstrukturiertheit vor. Hierdurch erhalten Stammdaten neben ihrer Größe eine weitere Big-Data-relevante Dimension, nämlich die Strukturiert- bzw. Unstrukturiertheit der Datensätze. Damit könnte man viele Stammdaten bereits als „Big Data“ bezeichnen.

5.1 Big Data

Um das Big-Data-Konzept zu verstehen, ist es hilfreich, sich den Vorgang einer Informationsanfrage vor Augen zu führen. Zu Beginn der Anfrage müssen Daten erhoben und gesammelt werden, aus denen später die Informationen gewonnen werden sollen. Wem hier der Unterschied zwischen Daten und Informationen nicht klar ist, dem könnte dieses anschauliche Beispiel Klarheit verschaffen:

Ein QR-Code enthält Informationen, wie beispielsweise einen Link zu einer Internetseite. Diese Information ist für den Betrachter ohne Hilfsmittel allerdings nicht verständlich. Erst durch einen QR-Code-Reader, wie man ihn vom Smartphone kennt, wird das Datum – der QR-Code – analysiert und gibt seine Information – die URL, also den Link zur Internetseite – dem Benutzer preis. Das bedeutet folglich, dass Daten eine strukturierte, möglicherweise codierte Darstellung von Informationen sind.

Daten können an den verschiedensten Quellen generiert werden, zum Beispiel an Computern, im *World Wide Web*, beispielsweise bei einer Suchmaschinen-Suche oder in Sozialen Netzwerken, durch Kameras und Sensoren oder von eingebauten Microprozessoren, die sich heutzutage bereits in vielen Geräten befinden. Natürlich gibt es hier noch viele weitere Datenquellen. Diese müssen auch nicht zwingend digital sein. Wir selbst sind auch große Datenlieferanten. Jede Aktion kann als Datenpaket aufgefasst werden, seien es Bewegungsdaten, Entscheidungsdaten oder Gedanken und Worte. Diese sind zwar in einem Big-Data-System schwieriger zu handhaben, da diese Daten erst digitalisiert werden müssen, doch auch hierzu gibt es bereits Lösungen. Fitnessarmbänder beispielsweise zeigen dem Läufer nach dem Training seine zurückgelegte Strecke und Vitalwerte an. Durch das Fortschreiten der Digitalisierung können somit immer mehr Daten digital zur Verfügung gestellt werden, mit denen immer bessere und komplexere Analysen durchgeführt werden können.

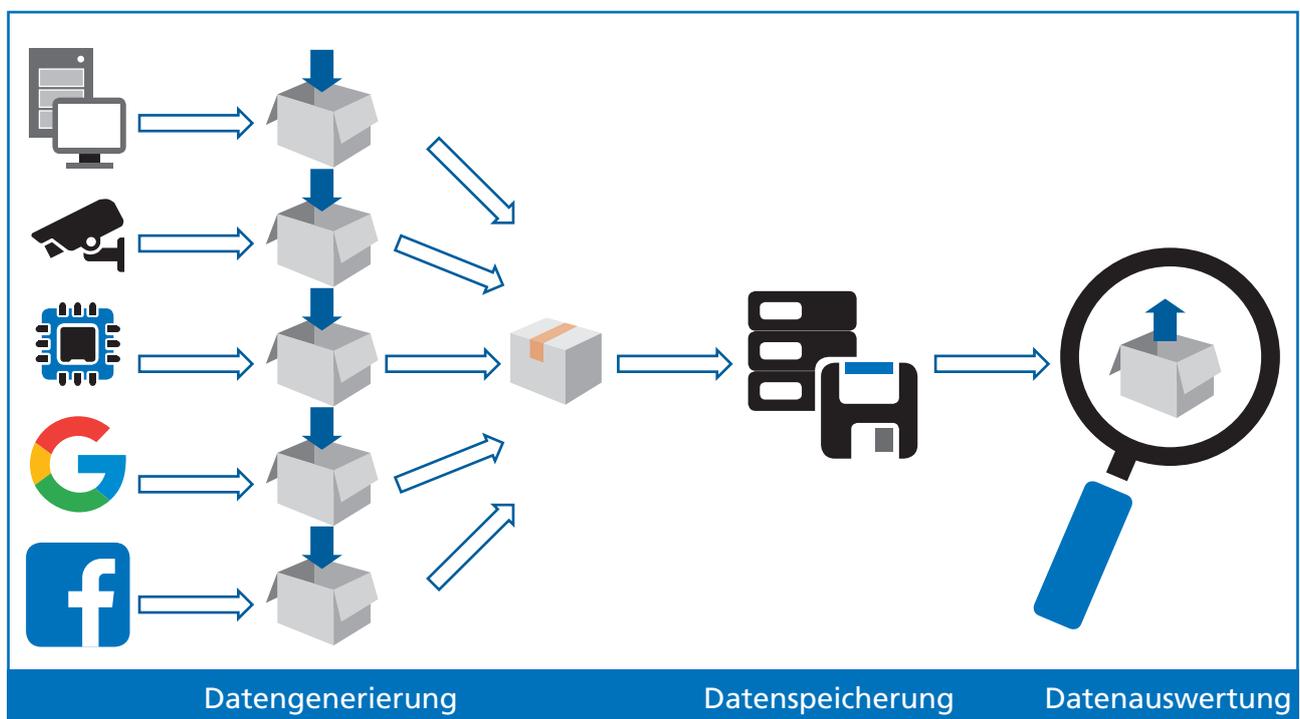


Bild 11: Big-Data-Konzept (eigene Darstellung)

Einmal erhoben, werden diese Datenpakete vor der späteren Analyse in einem Datenspeicher gespeichert. Hierfür gibt es je nach System unterschiedliche Möglichkeiten. Die klassische Herangehensweise wäre die Speicherung in einer Datenbank, die nach Abschluss der Recherche mit entsprechenden Suchanfragen nach Zusammenhängen durchsucht wird. Andere Techniken nutzen den Arbeitsspeicher, da eine langfristige Sicherung der Daten vielleicht nicht relevant ist oder erst nach erfolgreicher Suche durchgeführt wird, um das Datenaufkommen zu minimieren. Vorteil dieser Methode ist es, dass auf im Arbeitsspeicher liegende Datensätze schneller zugegriffen werden kann.

Nach der Speicherung werden die Datenpakete mithilfe von Big-Data-Methoden, also Algorithmen, die für die Suche in großen Datenmengen optimiert wurden, wobei hier die Größe wieder relativ ist, durchsucht. In Bild 11 wird die Analyse durch ein Öffnen eines Pakets symbolisiert. Dies soll verdeutlichen, dass die gewünschten Informationen aus den Datenpaketen extrahiert werden. Wie man sich nun leicht vorstellen kann, benötigt das Öffnen der Datenpakete – weiterhin gehen wir bei Big Data von einer großen Datenmenge aus – dementsprechend viel Zeit. Das Speichern und insbesondere das Öffnen, also die Extraktion der Information aus den Daten und die Verarbeitung der so gewonnenen Informationen, aller Pakete sind in diesem Konzept die limitierenden Faktoren, die eine Echtzeitanalyse erschweren. Die Geschwindigkeit solcher Algorithmen wird zwar stetig verbessert; auch die Hardware, die diese Algorithmen ausführt, wird immer schneller, dennoch werden sie nicht mit der explosionsartigen Steigerung der Datenmengen mithalten können. Hier kommt das Big-Data-Konzept an seine Grenzen. Ein alternatives Konzept wird im folgenden Abschnitt vorgestellt und dient zum Vergleich.

5.2 Smart Data

Das Smart-Data-Konzept ist grundsätzlich ähnlich dem Big-Data-Konzept aufgebaut. Die drei Aufteilungen des Konzepts sind auch in Smart Data in nahezu gleicher Weise vertreten. Sie werden lediglich an das Smart-Data-Konzept angepasst. Beginnt man wiederum mit der Datengenerierung, so bleibt dieser Prozess identisch. Es werden die gleichen Daten wie bei einer Big-Data-Analyse erhoben, die Menge wird also nicht zwangsläufig reduziert. Man könnte beispielsweise Daten aus sozialen Netzwerken vernachlässigen, wenn man nach Störungsmustern in der Produktion sucht, da diese höchstwahrscheinlich keine relevanten Informationen hierzu beitragen, aber hier liegt nicht das Hauptaugenmerk im Unterschied zu Big Data.

Im nächsten Schritt kommt nun die Änderung von Smart Data im Vergleich zu Big Data. Bevor man die Daten in einem Speicher ablegt, werden die Datenpakete mit Metainformationen angereichert. Diese Metainformationen – in Bild 12 mit Barcodes dargestellt – beinhalten weitere Informationen wie zum Beispiel die Herkunft der Daten, einen Zeitstempel oder den Zusammenhang, aus welchem sie stammen. Bei der Produktion wäre dies beispielsweise der Maschinenzustand, also ob es sich bei dem Datenpaket um Regeldaten oder Stördaten handelt. Durch diese Datenkategorisierung wird die spätere Analyse vereinfacht und somit beschleunigt.

Nach der Kategorisierung werden die Daten analog zum Big-Data-Konzept in einem Speicher hinterlegt. Dies geschieht jedoch nun durch die Kategorisierung der Daten in dafür vorgesehenen Kategorien. Stördaten werden zum Beispiel getrennt von Regeldaten gespeichert, um einen schnelleren Zugriff auf bestimmte Datenpakete zu erhalten. Der Vorteil dieser Unterteilung der Datenpakete ist es, dass nur die für eine Anfrage relevanten Datenpakete untersucht werden können, ohne dass man diese erst lange in der Datenbank suchen muss. Daten, die hierzu keine Informationen beitragen können, werden vor der Analyse bereits aussortiert und müssen nicht mehr „geöffnet“ werden.

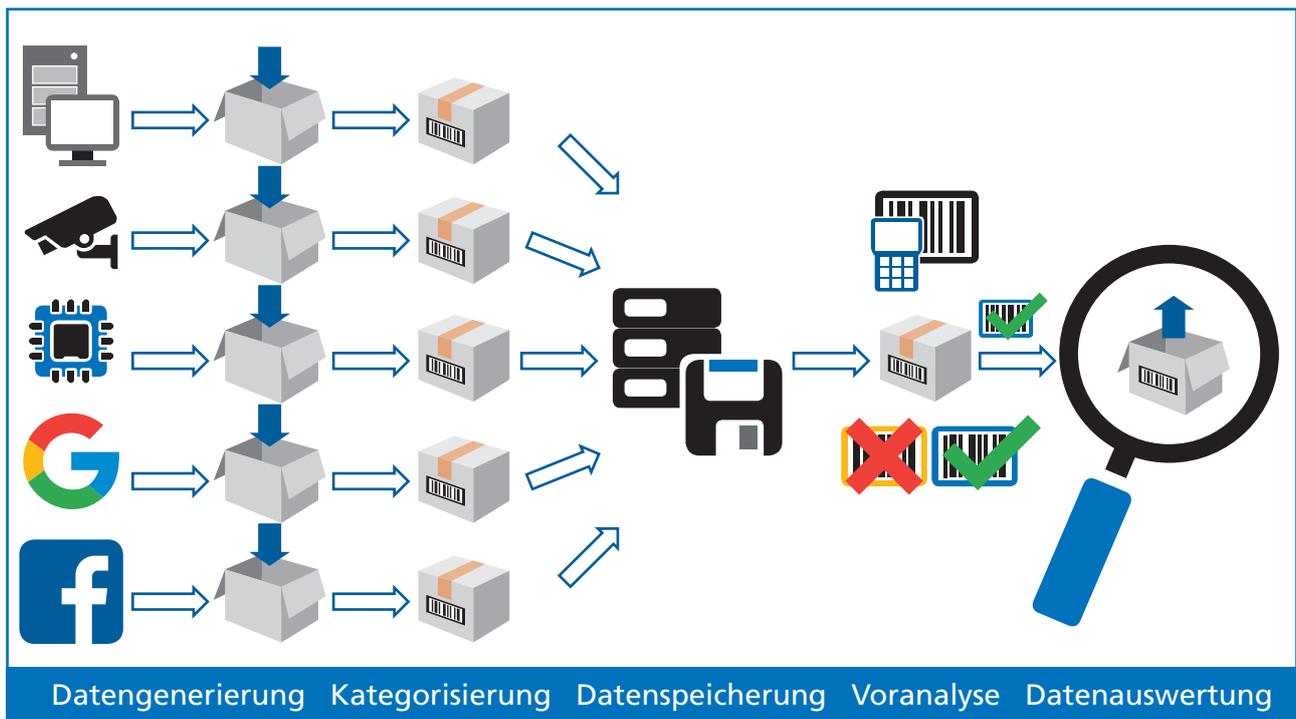


Bild 12: Smart-Data-Konzept (eigene Darstellung)

Die Datenanalyse wird im Smart-Data-Konzept also dadurch beschleunigt, dass nur noch die für die Anfrage relevanten Daten ausgewertet werden müssen. Dies lässt sich durch die Vorsortierung in die entsprechenden Kategorien leicht realisieren. Jedoch ist hier Fachwissen vonnöten, welches vorab entscheidet, welche Daten für die Anfrage relevant sind und analysiert werden müssen. Dieses Fachwissen geht als die zuvor beschriebenen Metainformationen in die Datenpakete ein und liefert so weitere Entscheidungskriterien bezüglich ihrer Relevanz zu bestimmten Fragestellungen. Das heißt also, dass Smart Data im Gegensatz zu Big Data auf Fachwissen in Form von Metainformationen basiert, welches bei der Analyse der Daten einbezogen wird. Dieses Fachwissen beschleunigt die Datenanalyse am Schluss, da die Menge der Datenpakete geringer ist und somit die

Analyse schneller durchgeführt werden kann. Allerdings basieren diese Analysen dann auf dem Fachwissen, können also nur bereits bekannte Zusammenhänge von vordefinierten Modellen erkennen. Bislang unbekannte Zusammenhänge werden gegebenenfalls nicht erkannt, da Datenpakete als irrelevant eingestuft wurden und nicht betrachtet wurden.

Dieser Vorteil ist auch gleichzeitig ein Nachteil: Während Smart-Data-Analyseverfahren auf Fachwissen aufbauen und hier sofort kategorisiert werden, wird in Big-Data-Verfahren zuerst ausschließlich, ohne weitere Kategorisierung, gesammelt, gleichsam Datenbestand angehäuft. Dieser vermeintliche Nachteil dient am Ende der breiteren Analysemöglichkeit, da mittels Big Data Zusammenhänge erkannt werden, die bei Vorkategorisierung gar nicht auffallen würden.

5.3 Smart Information

Die Konzepte Big Data und Smart Data liefern dem Nutzer auf Anfrage Informationen. Dies ist aber nur der erste Schritt in Zeiten von Industrie 4.0 und Smart Factory. Wäre es nicht schöner, wenn diese Anfragen nicht erst vom Nutzer gestellt werden müssten und er die gewünschten Informationen quasi in Echtzeit erhalten würde? Die Idee des Smart-Information-Konzepts ist es, dass die Informationen bereits so „smart“ sind, dass sie sich über das System selbst an die zuständigen Nutzer weiterleiten und neben ihrem Informationsgehalt auch noch weitere Informationen zur Entscheidungsunterstützung und zur Umsetzung der Entscheidung mitliefern.

Auch dieses Konzept lässt sich besser an einem Beispiel veranschaulichen:

In einem produzierenden Betrieb tritt bei einer Maschine eine Störung auf, die die Produktion zum Erliegen bringt. Nun erkennt das System diese Störung und übermittelt diese Information dem Wartungsmechaniker inklusive der Informationen zu den benötigten Ersatzteilen und einer Anleitung zur Reparatur. Gleichzeitig informiert das System den Produktionsleiter über den Ausfall und die Auswirkungen, dass beispielsweise der Liefertermin der aktuellen Produktserie mit der aktuellen Produktionsplanung nicht mehr eingehalten werden kann. Neben der reinen Information „Liefertermin kann nicht eingehalten werden“ werden ihm bereits mögliche Entscheidungsszenarien mitgeliefert. Dies können verschiedene Szenarien von der Änderung der Produktionsplanung über Informationen zur Bestellung von Produkten externer Fabrikanten bis hin zu Kontaktinformationen des Kunden sein, um ihn um Aufschub für die Lieferung zu bitten.

Hierdurch hat der Anwender den Vorteil, dass er die benötigten Informationen zum weiteren Vorgehen nicht erst noch suchen muss, sondern direkt mitgeliefert bekommt. Dies macht die Information „Liefertermin kann nicht eingehalten werden“ durch die Bereitstellung nützlicher und relevanter Informationen zu einer „smarten“ Information.

Dieses Konzept baut auf einer im Hintergrund laufenden Smart-Data-Analyse auf, die die entsprechenden Szenarien anhand vorgegebener

Modelle berechnet und auswertet. Hierzu muss zunächst eine große Wissensdatenbank angelegt werden, die das Prozesswissen, das Produktwissen, das Maschinenwissen und die Kundendaten, also die Stammdaten des Unternehmens, vereint und die von den Maschinen und möglicherweise auch Mitarbeitern ausgehenden Daten mit eben diesen Metainformationen anreichert, damit die Analyse hieraus in Echtzeit die „smarten“ Informationen zusammenstellen und entsprechend verteilen kann.

5.4 Beispiele und Anwendungsfälle

Wie bereits im vorangegangenen Abschnitt erwähnt, erforschen wir am FIR zusammen mit unseren Partnern unter anderem, wie anhand von Daten und Datenmustern Störungen prognostiziert und proaktiv gemanagt werden können. Im Rahmen des Forschungsprojekts BigPro werden (neue) Verfahren entwickelt, Daten aus der Produktion digital zu erfassen. Diese Daten werden analysiert und auftretende Muster werden im System hinterlegt. Anhand von Live-Daten, die mit Big-Data-Methoden mit den bekannten Mustern abgeglichen werden, wird die Auftrittswahrscheinlichkeit einer Störung berechnet und in einem späteren Schritt die bestmögliche Reaktion hierauf ermittelt. Dieses Vorgehen bezeichnen wir als *proaktives Störungsmanagement*.

Um dieses Störungsmanagement umzusetzen, mussten teils neue Konzepte zur Datenerfassung entwickelt werden. Hierbei wurden die hohen Temperaturen in der Stahlgießerei – eines unserer Partnerunternehmen – für die Standard-RFID-Tags zum Verhängnis, in der DFA (Demonstrationsfabrik Aachen) stellte uns die digitale Datenerfassung der manuellen Montage vor eine schwierige Aufgabe. Wie bereits erwähnt, produziert jeder von uns durch seine Handlungen Daten, doch ist die digitale Erfassung solcher Daten nicht immer einfach. In der DFA wurden daher die Anleitungen digitalisiert, d. h., statt in einer Anleitung aus Papier zu blättern, sind nun an den Arbeitsplätzen Tablets installiert, auf denen eine digitale Anleitung aufgespielt ist. Anhand des Umblätternvorgangs (Tastendruck) können nun Prozesszeiten digital erfasst werden. Die hierdurch erfassten Daten fließen neben vielen anderen Daten nun mit in die Analyse ein und verschärfen hierdurch die Störungsprognostik.

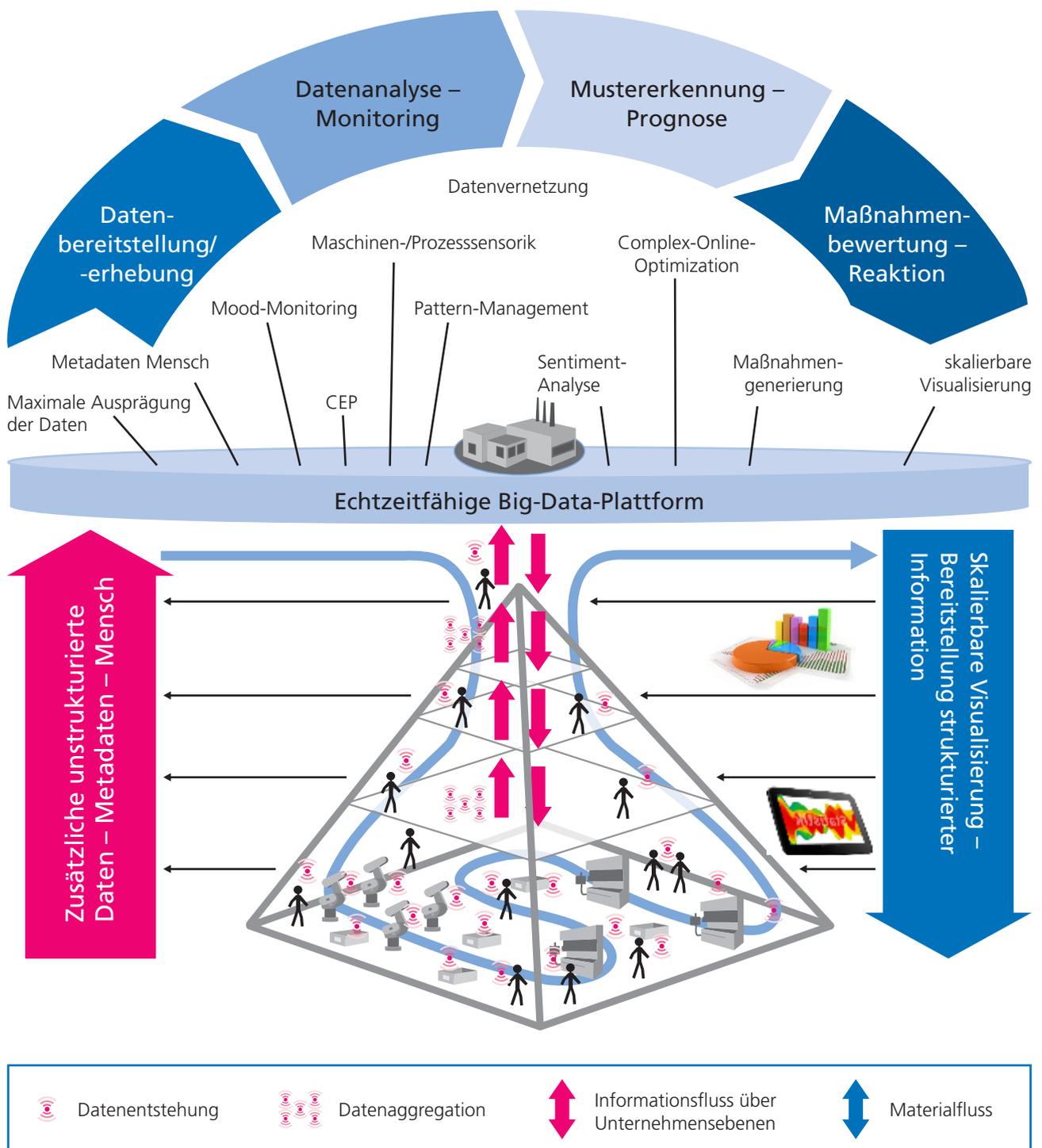


Bild 13: Zielbild BigPro (eigene Darstellung)

Diese gesamte Methodik funktioniert allerdings nur, wenn man die dort erfassten Daten mit den Stammdaten des Unternehmens, beispielsweise mit den Maschinendaten, den Auftragsdaten und den Materialdaten, verknüpft. Ohne die Stammdaten können Maschinenzustände zwar analysiert und Maschinenverhalten vorhergesagt

werden, es können jedoch nicht proaktiv Entscheidungshilfen erstellt werden. Stammdaten spielen also im proaktiven Störungsmanagement eine entscheidende Rolle. Daher gilt auch hier: Gepflegte Daten beschleunigen die Analyse und verbessern somit die Ergebnisse. Stammdatenmanagement lohnt sich also!

6 Das FIR als kompetenter Partner in der Praxis

Das FIR ist eine gemeinnützige, branchenübergreifende Forschungseinrichtung an der RWTH Aachen auf dem Gebiet der Betriebsorganisation und Unternehmensentwicklung mit dem Ziel, die organisationalen Grundlagen zu schaffen für das digital vernetzte industrielle Unternehmen der Zukunft.

Das Institut begleitet Unternehmen, forscht, qualifiziert und lehrt in den Bereichen Dienstleistungsmanagement, Business-Transformation, Informationsmanagement und Produktionsmanagement. Als Mitglied der Arbeitsgemeinschaft industrieller Forschungsvereinigungen fördert das FIR die Forschung und Entwicklung zugunsten kleiner, mittlerer und großer Unternehmen. Seit 2010 leitet der Geschäftsführer des FIR, Professor Volker Stich, zudem das Cluster Smart Logistik auf dem RWTH Aachen Campus.

Das Cluster Smart Logistik ist eines der sechs Startcluster auf dem Campus Melaten. Über 350 Menschen aus Wissenschaft und Wirtschaft erforschen und entwickeln dort Lösungen, wie Waren und Informationen in einer digitalen Welt der Zukunft optimiert vernetzt werden können.

Ausgerichtet auf eine völlig neue Form der Zusammenarbeit zwischen Forschung und Industrie werden die komplexen Zusammenhänge in realen Produktions- und IT-Umgebungen erlebbar gemacht. Neben den Innovation-Labs, in denen der inner- und überbetriebliche Daten- und Informationsaustausch live simuliert werden, bietet die angeschlossene Demonstrationsfabrik die Möglichkeit, aktuelle Forschungsergebnisse in einer echten Produktion zu erproben und zu veranschaulichen.

In diesem Zusammenhang verfügt das FIR über langjährige Erfahrung im Bereich Stammdatenmanagement, sowohl durch Forschungsprojekte als auch durch Beratung der Praxis. Durch das vom FIR entwickelte, methodisch fundierte Vorgehen zur Stammdatenoptimierung wurden in zahlreichen Projekten die Datenbestände verschiedener Unternehmen strukturiert, harmonisiert und die Datenpflege nachhaltig organisiert.



7 Literaturverzeichnis

HAUG, A.; STENTOFT ARLBJØRN, J.; ZACHARIASSEN, F.; SCHLICHTER, J.: „Master data quality barriers: an empirical investigation“. In: *Industrial Management & Data Systems* 113(2013)2, S. 234 - 249.

HEILER, R. J. (Hrsg.): E-Procurement. Effiziente Prozesse und Kostentransparenz im Einkauf. <http://docplayer.org/4212847-E-procurement-effiziente-prozesse-und-kostentransparenz-im-einkauf.html> (letzter Zugriff: 17.01.2017)

KHAN, Q.: Impact of RACI on Delivery & Outcome of Software Development Projects, S. 178. <http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=6783449> (zuletzt geprüft: 17.03.2017)

LOSER, C.: Master Data Management For Collaborative Service Processes. Institute of Information Management, University of St.Gallen, Switzerland, 2008.

OTTO, B.; HÜNER, K. M.; ÖSTERLE, H.: Unternehmensweite Stammdatenqualität. In: *ERP-Management* (2009)3, S. 19-21.

SCHEIBMAYER, M.; KNAPP, M.: [Studie] Stammdatenmanagement in der produzierenden Industrie. Hrsg.: G. Schuh. FIR e. V. an der RWTH Aachen, 2014.

WEISS, H.: Big Data war gestern – Huge Data ist heute. Mehr Ernüchterung als Euphorie. In: *Computerwoche online*, 14.10.2015. <http://www.computerwoche.de/a/big-data-war-gestern-huge-data-ist-heute,3217219> (zuletzt geprüft: 17.03.2017)

Haben Sie Interesse an weiteren Informationen aus unserem Haus?

Dann melden Sie sich gerne zu einem oder mehreren themenspezifischen Newslettern an:



newsletter-anmeldung-wp.fir.de