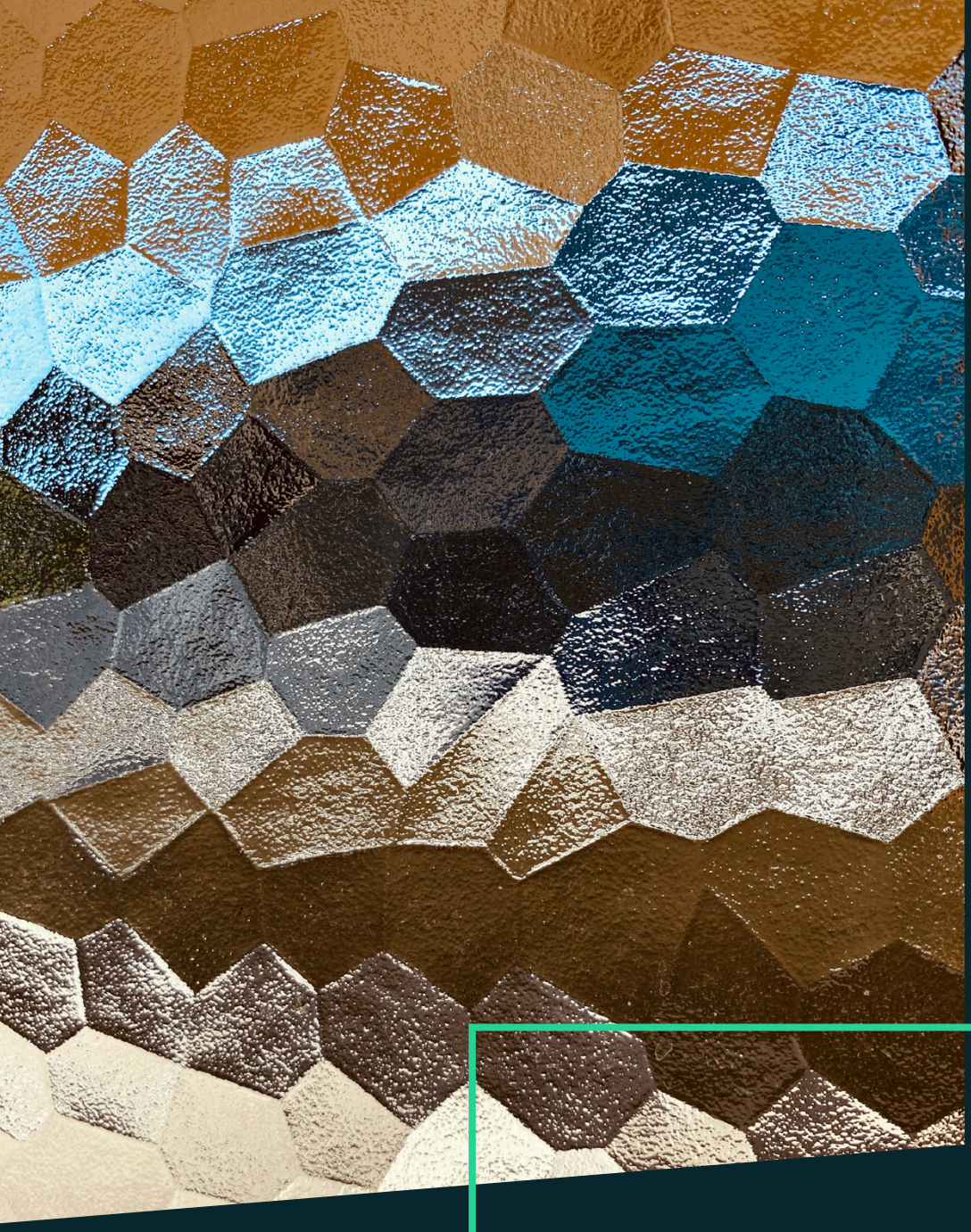


bitkom

Best Practices zur Entwicklung von Datenprodukten

Herausgeber
Bitkom e. V.
Albrechtstraße 10 | 10117 Berlin

Ansprechpartner
David Schönwerth | Bitkom e. V.
Bereichsleiter Data Economy
T 030 27576-179 | d.schoenwerth@bitkom.org

Autorinnen und Autoren
Marcel Altendeitering (Fraunhofer ISST), Adel Anwar (PwC Deutschland), Stephan Bautz (PwC Deutschland), Dr. Marcel-Philippe Breuer (DB Systel GmbH), Chris Buchhold (Deutsche Bahn AG), Lukas Feuerstein (Deloitte Consulting GmbH), Marcus Hartmann (PwC Deutschland), Pascal Hess (Fujitsu Services GmbH), Dr.-Ing. Ibrahim Halfaoui (TÜV SÜD Digital Service GmbH), Dr. Marvin Jagals (Deutsche Bahn AG), Ralph Kemperdick (RaKeTe-Technology), Dr. Michael Kraus (CMS Hasche Sigle), Andreas Kulpa (CRIF GmbH), Prof. Dr. Chris S. Langdon (T-Systems International GmbH), Dr. Till Luhmann (BTC Business Technology Consulting AG), Dr. Michael Pauly (Deutsche Telekom Geschäftskunden GmbH), Nina Popanton (T-Systems International GmbH), Dr. Daniel Pöppelmann (DB Systel GmbH), Felix Ruscheweyh (Swisslog GmbH), Dr. Kerstin Schmidt (CRIF GmbH), David Schönwerth (Bitkom e. V.), Theresa Schramm (Deutsche Telekom Geschäftskunden GmbH), Marc Schumacher (GP+S Consulting GmbH), Dr. Michael Stadler (BTC Business Technology Consulting AG), Thorn Thaler (CRIF GmbH), Dr. Adam Trendowicz (Fraunhofer IESE), Umair Usman (PwC Deutschland), Dr.-Ing. Sebastian Werner (Thoughtworks Deutschland GmbH)

Verantwortliche Bitkom-Gremien
AK Data Strategy & Data Products

Layout
Sabrina Flemming | Bitkom e. V.

Titelbild
© susan gold – unsplash.com

1 Einführung: Daten im Unternehmen nutzen 6

2 Entwicklungsprozess: In 6 Stufen zum Datenprodukt 10

Discover: Potenzial und Einsatzmöglichkeiten 12

Define: Bewertung und Priorisierung der Handlungsoptionen 13

Design: Konzeptionelle Entwicklung konkreter Datenproduktideen 14

Deliver: Entwicklung des Datenprodukts 14

Operate: Betrieb des Datenprodukts 15

Retire: Archivierung oder Löschung von Datenprodukten 16

3 Governance: Entwicklung von Datenprodukten 17

Grundlagen: Worauf zu achten ist 18

Einführung eines unternehmensweiten Datenkataloges 19

Wechselwirkung zwischen Datenkatalog und Datenprodukt 21

Datenprodukte im Datenkatalog 22

Data Marketplace: mit Daten handeln 25

Automatisierungsansätze für Data Governance und Datenqualität 26

4 Datenqualität steigern und absichern 28

Qualität: Aspekte, Merkmale, Attribute 29

Qualitätsmodell 31

Qualitätsmanagement für Datenprodukte 33

Toolgestütztes Qualitätsmanagement 35

5 Rechtliche Aspekte 36

Kein Dateneigentum 37

Datenschutz 38

Geschäftsgeheimnisse 39

Weitere geschützte Inhalte 39

Datenverträge 40

Fazit 43

6 Zusammenfassung 44

7 Praxisbeispiele 46

Einsatzbeispiel Deutsche Telekom 47

Einsatzbeispiel Swisslog / GP+S Consulting 49

Einsatzbeispiel PwC Deutschland 51

Einsatzbeispiel Deutsche Bahn 53

Einsatzbeispiel CRIF 55

8 Mitwirkende 57

Abbildungen

1 Vorgehensmodell zum Datenprodukt (verändert nach GP+S Consulting, ursprünglich British Design Council) 11

2 Ausprägungen eines Metadatenstandards für Datenprodukte 22

3 Möglichkeit der Qualitätsbewertung 30

4 Grundbestandteile eines Datenqualitätsmodells 31

5 Beispiel für ein einfaches Datenqualitätsmodell 31

6 Übersicht für einen Datenqualitätsprozess auf Basis von Wang, R. Y. (1998) 34

7 Übersicht Data Sharing Prozesse auf Basis von Altendeitering, M et al. (2022) 35

8 App-Architektur: Intermodales Reisen in Hamburg 48

1 Einführung: Daten im Unternehmen nutzen

1 Einführung: Daten im Unternehmen nutzen

Anlass

Lesende dieses Leitfadens sind höchstwahrscheinlich bereits vom Potenzial von Daten für den Unternehmenserfolg überzeugt. Trotzdem sieht die Realität in Unternehmen oft anders aus. Daten fragwürdiger Qualität liegen in Silos, das Management hat Wichtigeres zu tun und eine systematische Verwertung findet nicht statt. Seit Jahren wird eine Diskussion um Datenformate sowie Datenkultur geführt und jede neue Task-Force dazu übergibt nach Auflösung den Staffelstab an die darauffolgende.

Erfolgreiche Geschäftsmodelle stützen sich zunehmend auf Daten, um innovative Produkte und Dienstleistungen zu schaffen. Heutzutage wird in diesem Kontext oft von einem Datenprodukt gesprochen:

> Daten, die für ihre Nutzer einen Mehrwert bringen und das Ergebnis eines »Produktionsprozesses« sind.

Folglich werden Datenprodukte vermehrt zu einem entscheidenden Unternehmenswert, der einen erheblichen Geschäftswert darstellt.

Dieser Leitfaden beantwortet nicht die Frage, welche Rolle Daten im Unternehmen spielen sollten, sondern gibt Anregungen, wie Datenprodukte effizient und systematisch entwickelt werden können.

Dieser Leitfaden richtet sich gezielt an Unternehmen, die mit ihren Daten Wertschöpfung oder Kostenersparnis erzielen möchten. Er bietet praktische Unterstützung und konkrete Handlungsoptionen, um die Umsetzung dieser Ziele erfolgreich zu gestalten.

Teils gibt es im Unternehmen bereits eine Datenstrategie, welche die Entwicklung von Datenprodukten leitet. Teils ist das nicht der Fall, dann sind Datenprodukte nicht nur das Ziel, sondern auch der Weg. Mit vorzeigbaren, überzeugenden Datenprodukten (einem Prototyp) lassen sich Unterstützung und Ressourcen gewinnen, welche helfen, Datenprodukte in der nächsten Strategiediskussion zu demonstrieren und so zu rechtfertigen.

Tritt man einen Schritt zurück, stellen sich größere strategische Fragen, bei denen die Antwort nicht immer einfach ist.

- Welchen Gestaltungsansatz verfolgen wir? Architecture-First oder Product-First?
- Wird die Datennutzung zentral geplant und ausgeführt (top-down) oder werden dezentral Ideen pilotiert und umgesetzt (bottom-up)?
- Organisieren sich Teams und Abteilungen in eigenen Entwicklungsprozessen oder wird dies von oben vorgegeben?
- Was lässt sich automatisieren? Was gibt es rechtlich zu beachten?

Definition von Datenprodukten

Laut Übersichten in der Literatur werden heute mehr als 80 Prozent des Zeitaufwands für Datenanalyse für die Datenaufbereitung aufgewendet – nicht etwa für die Entwicklung von Algorithmen.[1] Dies würde das 80/20 Pareto-Prinzip, einen Eckpfeiler der Geschäftseffizienz, auf den Kopf stellen.[2] Die Lösung: Daten wie Produkte behandeln und bewährte Ansätze, wie das Produktmanagement auf Daten anwenden.[3] Es geht darum, die Daten als eigenständiges Produkt zu verstehen und nicht als Nebenprodukt einer anderen Tätigkeit. Zu einem Datenprodukt gehört alles, das nötig ist, um dieses bereitzustellen: Daten, Metadaten, Code, Policies und ggf. Infrastruktur. Das Produkt muss autonom und damit eigenständig nutzbar sowie vom Anbieter mit einem Service Level versehen sein, welches eine verlässliche Nutzung garantiert. Auf Führungsebene werden »Daten« tendenziell als Endprodukt angesehen, das man »von der Stange« kaufen kann. Für Datenwissenschaftlerinnen und Datenwissenschaftler ähneln sie jedoch eher einem Inputfaktor. Dieser muss aufbereitet / weiterverarbeitet werden – es muss also Wertschöpfung stattfinden, etwa durch Formatieren, Korrigieren, Beschriften, Analysieren. Dies erfordert Ressourcen und erklärt, warum etwa die Lizenzierung von Datensätzen oder Datenprodukten mit bestimmten Eigenschaften signifikant höhere Preise erzielen kann als der Zugriff auf Rohdaten. Bei Daten fehlen heute oft Beschreibungen zu (a) Informationsgehalt, (b) Qualitätsbeurteilung und (c) Mengenindikation.[4]

Ein Produkt sollte diese Eigenschaften erfüllen: es muss auffindbar (»discoverable«), verständlich (»understandable«), adressierbar (»addressable«), sicher (»secure«), interoperabel (»interoperable & composable«), vertrauenswürdig (»trustworthy«), einfach und nativ abrufbar (»natively accessible«) und eigenständig nutzbar (»valuable on its own«) sein.[5]

1 Press, G. 2016. Cleaning Big Data: Most Time-Consuming, Least Enjoyable Data Science Task, Survey Says. Forbes (March 23rd), Vollenweider, M. 2016. Mind+Machine: A Decision Model for Optimization and Implementing Analytics. John Wiley & Sons: Hoboken, NJ

2 Z.B. Newman, M.E. 2005. Power laws, Pareto Distributions, and Zipf's law. Contemporary Physics 46(5): 323–351

3 Crosby, L., and C. Schlueter Langdon. 2019. Data as a Product to be Managed. Marketing News, American Marketing Association (April 24th), ↗https://www.ama.org/marketing-news/data-is-a-product

4 Schlueter Langdon, C., Sikora, R. 2020. Creating a Data Factory for Data Products. In: Lang, K.R., et al. Smart Business: Technology and Data Enabled Innovative Business Models and Practices. WeB 2019. Lecture Notes in Business Information Processing, vol 403. Springer, Cham. ↗https://doi.org/10.1007/978-3-030-67781-7_5

5 Bitkom. 2022. Data Mesh – Datenpotenziale finden und nutzen. ↗https://www.bitkom.org/Bitkom/Publikationen/Data-Mesh-Datenpotenziale-finden-und-nutzen

Erfahrungen bei geringer Datenqualität

Egal, ob Daten im Rohformat, als Datenprodukt oder in einem Zwischenstadium vorliegen, hat eine schlechte Datenqualität neben Aufwänden zur Bereinigung auch Auswirkungen auf das »echte« Leben.

Auswirkungen im Unternehmen können z. B. die Folgenden sein:

- Umsatzverlust – Ist ein Artikel in einem Onlineshop einer falschen Kategorie (male statt female) zugeordnet, wird er nicht gefunden und damit nicht gekauft.
- Verärgerte Kunden – Hat man die falsche Adresse einer Filiale auf der Webseite, verärgert man Kunden, die zu einem kommen möchten.
- Rechtsrisiken – Veraltete Informationen führen ggf. zu nicht (mehr) erlaubten Entscheidungen.
- Insolvenz – Unternehmerische Entscheidungen basierend auf falschen Daten können zum Untergang eines ganzen Unternehmens führen.

Doch auch konkrete Situationen in der Geschichte haben mehr oder weniger große Auswirkungen auf das Leben der Menschen:

- Am 23.09.1999 scheiterte die Marsmission des »Mars Climate Orbiters«, da die Bodenkontrolle die Daten in Poundforce sendete, aber der Orbiter die Daten in SI-Einheiten interpretierte.[6]
- In einem Projekt wurden Daten bereinigt, indem die Werte »Hund« und »Pferd« aus dem Feld »Anrede« einer Mitarbeiterdatenbank entfernt wurden. Daraufhin scheiterten viele monatliche Zahlungsläufe. Die »Bereinigung« war falsch, da die Einträge richtigerweise zu Polizeihunden und Polizeipferden gehörten, denen ebenfalls Bezüge zugeordnet waren.
- Bei Patienten in Krankenhäusern werden (hin und wieder) in OPs die Seiten vertauscht und z. B. das rechte statt des linken Beins entfernt, weil es falsch vermerkt wurde.
- Bei Geschäften auf dem Finanzmarkt kommt es immer wieder zum »Fat-Finger-Fehler«, bei dem ein Tippfehler große Börsenbewegungen auslöst.[7]

6 Wunderlich-Pfeiffer, F. 2015. Softwarefehler in der Raumfahrt: In den Neunzigern stürzte alles ab. Golem.de. ↗https://www.golem.de/news/softwarefehler-in-der-raumfahrt-in-den-neunzigern-stuerzte-alles-ab-1511-117537-3.html

7 Wikipedia Deutschland. 2023. Fat-Finger-Fehler. ↗https://de.wikipedia.org/wiki/Fat-Finger-Fehler

2 Entwicklungsprozess: In 6 Stufen zum Datenprodukt

2 Entwicklungsprozess: In 6 Stufen zum Datenprodukt

Die wesentlichen Stakeholder eines Datenprodukts sind zum einen die Nutzenden, für welche das Produkt geschaffen wird und zum anderen die Entwicklerinnen und Entwickler, welche das Produkt bereitstellen sollen. Auf organisatorischer Ebene geht es dementsprechend einerseits um Kundinnen und Kunden, also Unternehmen oder Abteilungen, die das Produkt verwenden, und andererseits um Anbieterinnen und Anbietern, also Unternehmen oder Abteilungen, welche das Produkt zur Verfügung stellen. Bei der Entwicklung müssen unterschiedliche Perspektiven auf ein Datenprodukt einbezogen werden:

- Die Business-Perspektive, die darstellt, wie das Produkt wirtschaftlich ausgewertet wird
- Die fachliche Perspektive, die darstellt, in welchen Fachprozessen das Produkt genutzt wird und einen Kundennutzen entwickelt
- Die technische Perspektive, die darstellt, wie das Produkt umgesetzt wird

Auch in der Entwicklungsphase spielt die Einbindung der Kundensicht in Form eines direkten Customer-Feedbacks eine zentrale Rolle. Das hier abgebildete Vorgehensmodell umfasst sechs Schritte und basiert auf dem Double-Diamond-Prozess. Skizziert wird ein interaktives Vorgehen, welches sowohl die Entwicklerinnen und Entwickler als auch die späteren Nutzer des Datenprodukts einbezieht.

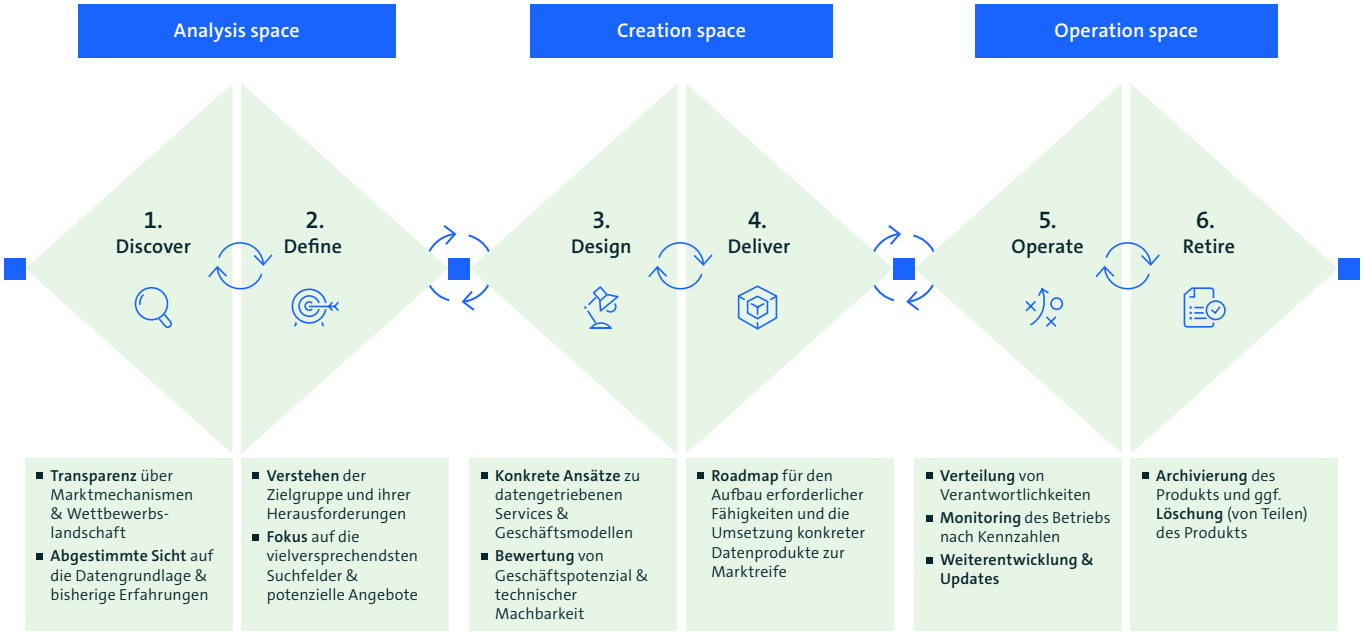


Abbildung 1 – Vorgehensmodell zum Datenprodukt (verändert nach GP+S Consulting, ursprünglich British Design Council)[8]

8 British Design Council. 2005. The Double Diamond: A universally accepted depiction of the design process. ↗https://www.designcouncil.org.uk/our-resources/the-double-diamond/. ↗CC-BY 4.0.

2.1 Discover: Potenzial und Einsatzmöglichkeiten

In der ersten Phase geht es darum, Potenzial und Einsatzmöglichkeiten für Datenprodukte zu erörtern. Es sollte, intern wie extern, einen konkreten (Informations-)Bedarf geben, der durch das neue Datenprodukt gedeckt wird. Alternativ können Vorgaben, z. B. aus einer Datenstrategie, Anlass sein, ein neues Datenprodukt zu entwickeln. Zudem sollten eine Erhebung der verfügbaren Daten sowie eine kritische Betrachtung der Fähigkeiten im Umgang mit Daten erfolgen. Um passende Datenprodukt-Ideen zu entwickeln, können Methoden wie beispielsweise das Business Model Canvas angewendet werden.

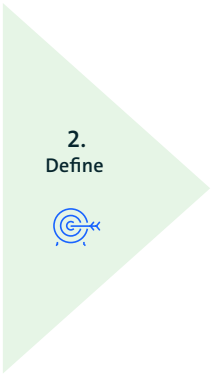
Ein konkreter Zweck von Datenprodukten besteht darin, Unternehmensdaten zu monetarisieren (in der externen Nutzung) und dabei mit einem bestimmten Geschäftsbereich zu beginnen. Dies wirft die Frage nach dem Markt und seinen Akteuren auf: Wer sind die zukünftigen Kundinnen und Kunden und wie sieht die Wettbewerbssituation aus? Kundinnen und Kunden können in der Innensicht beispielsweise Fachbereiche oder das Management sein, aber auch die Belegschaft insgesamt. Außerhalb der Organisation finden sich Kundinnen und Kunden in Form von Geschäftspartnern, Endkunden oder der Bevölkerung.

Bei der Analyse des Wettbewerbs besteht die Herausforderung, dass ein Teil des Marktes an Informationsangeboten in Unternehmen oftmals nicht offiziell erfasst wird (wie z. B. lokale Excel-Tabellen). Nicht nur die Analyse wird dadurch erschwert, auch im späteren Einsatz muss sich das Produkt gegen diese teils unbekannte Konkurrenz behaupten. Daher ist neben der Bedarfsermittlung auch zu klären, wie das Produkt seine Kundinnen und Kunden bestmöglich erreichen kann. Eine weitere Überlegung kann sein, inoffizielle Datenprodukte, wenn bekannt, zu formalisieren und zu standardisieren, was eine Neuentwicklung überflüssig macht.

Weil Datenprodukte für viele Unternehmen ein neues Betätigungsfeld darstellen, verfügen sie in diesem Bereich oftmals nur über wenige Kapazitäten und Erfahrungswerte. Aus diesem Grund ist es ratsam, zunächst Transparenz über die für Entwicklung und Betrieb notwendigen Fähigkeiten herzustellen. Dabei steht die Frage im Mittelpunkt, was das Unternehmen können muss, um die Datenproduktideen künftig erfolgreich realisieren, betreiben und skalieren zu können (technische und fachliche Capabilities). Mit der Methodik des Business Capability Mappings[9] lassen sich somit im Sinne einer Gap-Analyse, ausgehend vom Status Quo, konkrete Entwicklungsbedarfe identifizieren. Der Aufbau der Capabilities kann parallel zur Entwicklung eines ersten Datenprodukts erfolgen. Ausgewählte Aspekte dieses Vorgehens, z. B. Betrieb, Data Governance, Datenqualität sowie Datenkataloge werden in den folgenden Kapiteln betrachtet. Ist der Markt mit seinen Akteurinnen und Akteure abgegrenzt und untersucht, sowie erste Bedarfe ermittelt und die notwendigen Capabilities definiert, gilt es, im nächsten Schritt das weitere Vorgehen zum Aufbau von Datenprodukten zu definieren.

9 Wikipedia. 2023. Business Capability Model, ↗ https://en.wikipedia.org/wiki/Business_capability_model

2.2 Define: Bewertung und Priorisierung der Handlungsoptionen



Als Nächstes erfolgen die Bewertung und die Priorisierung der Handlungsoptionen: Welche Bedarfe oder Segmente sollten vorrangig bedient werden, welche Herausforderungen gilt es zuerst zu lösen?

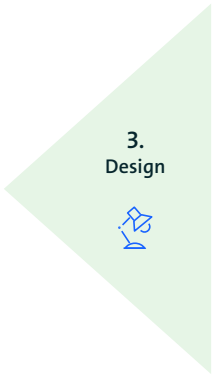
Die Wahl der Bewertungskriterien erfolgt auf individueller Basis. Daher soll an dieser Stelle nur eine Liste möglicher Optionen gezeigt werden:

- Strategischer bzw. unternehmerischer Fit, ggf. auch durch eine Entscheidung aus dem Management bzw. der Leitung
- Time to Market (wie lange benötigt es, diesen Bedarf decken zu können)
- Wirtschaftlichkeit (ist ein Vorgehen in diesem Markt gewinnbringend oder nutzenstiftend)
- Reifegrad des zum jeweiligen Bedarf bzw. der Herausforderung passenden Produktbereichs bzgl.
  - Daten (z. B. Qualität, Verfügbarkeit)
  - bestehender Capabilities (z. B. Betrieb und Support der Datenprodukte)
  - bestehender Ansätze (z. B. inoffizielle Datenprodukte)
- Geltungsbereich und Wiederverwendbarkeit/Perspektive (wie ist die Größe und lässt sich der Ansatz skalieren)
- Rechtliche, ethische oder moralische Vorgaben
- Wettbewerb

Diese Kriterien sind teilweise auch für die Bewertung von Ideen für konkrete Datenprodukte in der dritten Phase nutzbar.

Nach der Auswahl der Kriterien sollte eine geeignete Metrik entwickelt werden, um die Bedarfe und Herausforderungen basierend auf der Bewertung zu priorisieren. Eine Möglichkeit besteht darin, die Bewertung in Quadranten aufzuteilen, insbesondere, wenn es um relativ knappe Entscheidungen bei der Priorisierung geht. Im nächsten Schritt sollten Ideen für unterschiedliche Optionen entwickelt und diese anschließend in die Priorisierung einbezogen werden.

2.3 Design: Konzeptionelle Entwicklung konkreter Datenproduktideen



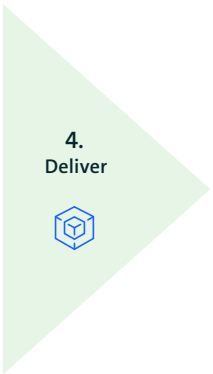
Im dritten Schritt des Vorgehens steht die konzeptionelle Entwicklung konkreter Datenproduktideen sowie eines ersten Prototyps im Mittelpunkt.

Wichtig ist, dass die Entwicklung kundenfokussiert erfolgt und auf die Lösung der in Schritt 2 priorisierten Herausforderungen abzielt.
Nach der initialen Identifikation möglicher Datenproduktideen ist es sinnvoll, die Ansätze anhand einiger Aspekte zu verifizieren, bevor eine detaillierte Ausarbeitung erfolgt. Ein Aspekt ist die Durchführung einer ersten groben Potenzialabschätzung, um den monetären Nutzen abschätzen und die einzelnen Produktideen miteinander vergleichen und priorisieren zu können.

Des Weiteren gilt es, die technische Machbarkeit der identifizierten Datenproduktideen zu überprüfen. Ein wichtiger Bestandteil ist hier die Prüfung der vorhandenen Datenbasis, inklusive der vorliegenden Datenqualität (↗ Kapitel 3), sowie die Anbindung der Systeme und Quellen. In diesem Zusammenhang sollte ein erster Prototyp im Rahmen eines PoC getestet werden.
Neben der Potenzialabschätzung und der technischen Machbarkeitsprüfung sind die rechtlichen Aspekte (↗ Kapitel 3.4) ein weiterer wichtiger Schritt. Vor der Umsetzung der Datenproduktideen ist zudem zu klären, wie das Produkt zukünftig vermarktet wird und zu welchem Preis es angeboten wird.

Es ist unerlässlich, die entwickelten und priorisierten Zielsetzungen mit den zukünftigen Anwendern zu diskutieren, um sicherzustellen, dass ihre Anforderungen frühzeitig berücksichtigt werden und ein nachträgliches Redesign vermieden werden kann.

2.4 Deliver: Entwicklung des Datenprodukts



In der Deliver-Phase wird die Organisation für den Betrieb und die Skalierung von Datenprodukten befähigt und aus dem Prototypen das finale Produkt entwickelt und ausgeliefert.

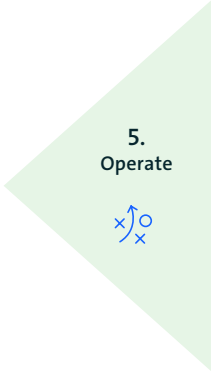
Mit der Entwicklung des ersten funktionalen Prototyps wurden in Phase 3 die konzeptionellen Datenproduktideen erstmals realisiert und die technische Machbarkeit bestätigt. Der Prototyp kann ersten (potenziellen) Kundinnen und Kunden präsentiert und von diesen getestet werden. Er dient somit zur strukturierten Aufnahme von Kundenfeedback und ermöglicht die unmittelbare Einbindung dieses Feedbacks in den weiteren Entwicklungsprozess.

Nachdem Transparenz über die erforderlichen Capabilities besteht, können diese in eine Umsetzungs-Roadmap überführt werden. Diese Umsetzungs-Roadmap stellt die

Capabilities in zeitliche Abfolge zueinander, legt Abhängigkeiten offen und ermöglicht einen fokussierten Aufbau der erfolgskritischen Capabilities.

Parallel dazu erfolgt der Aufbau der in Phase 1 ermittelten Capabilities für den Betrieb. Mit Hilfe des Kundenfeedbacks wird auf Basis des Prototyps das finale Datenprodukt entwickelt. Nach der abschließenden Qualitätssicherung wird es in den Betrieb übergeben.

2.5 Operate: Betrieb des Datenprodukts



Je nach Organisationsmodell, Reifegrad und technologischer Basis kann der Betrieb zentral, dezentral oder auch hybrid (z. B. durch eine zentrale Bewirtschaftung und ein dezentrales Monitoring) erfolgen. Zudem besteht eine Unterscheidung zwischen Fachbereichen, »Datenteams« und der IT, wobei auch die IT eigene Datenprodukte anbieten kann und nicht nur als Betreiber von Produkten anderer Bereiche infrage kommt. Aus der Vielzahl an Kombinationsmöglichkeiten ergibt sich keine klare Empfehlung, vielmehr sollte eine umfangreiche Abwägung erfolgen, wie stark welche Bereiche oder Teams Verantwortung für den Betrieb (sowie auch die Entwicklung) übernehmen können und sollen. Dafür kann die Beantwortung folgender Fragen hilfreich sein:

- Wie stark sind die Fachbereiche bereits durch das Tagesgeschäft ausgelastet?
- Welchen Mehrwert bietet eine Übernahme der Verantwortung durch die Fachbereiche (z. B. gestärktes Bewusstsein für den Wert von Daten, insbesondere im eigenen fachlichen Kontext, Aufbau von Kompetenzen etc.)?
- Ist eine Reifegrad-abhängige Verantwortung denkbar, bei der mit der Zeit von einem zentralen oder hybriden/CoE-Ansatz zu einem dezentralen Ansatz gewechselt wird?
- Können durch einen hybriden Ansatz Fachbereiche und IT stärker zusammengeführt werden?

Neben der Verantwortung sind zahlreiche Aktivitäten während des Betriebs durchzuführen. Grundsätzlich sollten alle Produkte einem Monitoring unterliegen, das Qualität (z. B. Konsistenz), Verfügbarkeit, Bedarf, Nutzung und auch die Kosten bzw. den Wert des Produkts überwacht. Im Idealfall werden Grenzwerte für Kennzahlen vorab definiert, um auf Veränderungen entsprechend reagieren zu können. Des Weiteren ist die notwendige Infrastruktur zu betreiben, die sich je nach Art des Produkts (z. B. Datensatz, Dashboard oder API) stark unterscheiden kann. Zudem sind Änderungen (z. B. als Releases) und Weiterentwicklungen durchzuführen, wie auch Marketingmaßnahmen oder Preisanpassungen.

2.6 Retire: Archivierung oder Löschung von Datenprodukten

Im Gegensatz zu der weit verbreiteten Annahme, dass Daten eine unbegrenzte Ressource sind, gilt auch und insbesondere bei Datenprodukten, dass die enthaltenen Informationen nur endlich nutzbar sind (da sich z. B. der Kontext ändert) und der zugrunde liegende Bedarf sich mit der Zeit verändern wird. Zudem bedeutet jedes Datum oder Datenprodukt Kosten in Lagerung, Nutzung, Betrieb und Verwaltung, denen ein entsprechender Wert bzw. Nutzen oder Zweck gegenüberstehen sollte, der sich mit der Zeit verändern wird.

Durch einen zweistufigen Retirement-Prozess, der durch Veränderungen im Betrieb des Produkts ausgelöst wird, wird diesen Tatsachen Rechnung getragen und sichergestellt, ein zu jeder Zeit allen Vorgaben entsprechendes Produkt zu betreiben.

Der erste Schritt besteht dabei in der Archivierung des Datenprodukts. Dies kommt dann zum Einsatz, wenn es beispielsweise gesetzliche Aufbewahrungsfristen notwendig machen. Sollte eine Archivierung nicht ausreichen, z. B. bei der Ausübung des Rechts auf Löschung von personenbezogenen Daten, wird das jeweilige Produkt (oder Teile davon) gelöscht.

3 Governance: Entwicklung von Datenprodukten

3 Governance: Entwicklung von Datenprodukten

3.1 Grundlagen: Worauf zu achten ist

Governance umfasst eine Vielzahl von Aspekten, insbesondere bei der Entwicklung von Datenprodukten ist es jedoch entscheidend, die richtigen Daten zur richtigen Zeit an die richtigen Personen zu liefern. Dies bedeutet, dass die folgenden Aspekte bei der Planung einer passenden Datenverwaltung und Data Governance[10] berücksichtigt werden müssen:

- Sicherheit: Sind die richtigen Personen authentifiziert und berechtigt, die Daten zu nutzen?
- Compliance: Entsprechen die Daten allen erforderlichen Richtlinien, z. B. DS-GVO, RTBF usw.?
- Verfügbarkeit: Sind die Daten für autorisierte Benutzende zugänglich?
- Qualität: Wie wird die Datenqualität in passender Weise quantifiziert und den Nutzenden mitgeteilt?
- Entitätsstandardisierung: Gibt es eine Einigung zur Terminologie in verschiedenen Bereichen?
- Provenance: Ist klar, wer für die Daten verantwortlich ist und woher sie stammen?

Bei jeder Art von Data Governance stellt sich die Frage, wie diese Governance durchgesetzt werden kann. Dies ist eine anspruchsvolle Aufgabe, und die Automatisierung der Durchsetzung der Konformität ist der notwendige Schlüssel dazu. Durch diese Automatisierung mit Vorschriften und Standards sollten Domänen standardmäßigen DevOps[11]-Praktiken folgen, um so ihre Governance anzuwenden.

Dies bedeutet, dass die Bedürfnisse und das spezifische Wissen der Domäne mit den übergreifenden Anforderungen des Unternehmens in Einklang gebracht werden müssen, um so eine klar formulierte und vereinbarte Aufteilung der Verantwortung und Zuständigkeiten zu erreichen.

Die Etablierung einer flexiblen und auf Wiederverwendung ausgerichteten Governance setzt eine partnerschaftliche und domänenübergreifende Zusammenarbeit voraus, eine weitere komplexe Herausforderung in Großunternehmen mit ihren vielfältigen Interessendivergenzen.

10 Wikipedia. 2023. Data Governance, ↗ https://en.wikipedia.org/wiki/Data_governance
11 Wikpedia. 2023. DevOps, ↗ https://en.wikipedia.org/wiki/DevOps

Ein effektiver Weg, dieser Herausforderung zu begegnen, ist es, jeweils eine Vertreterin oder einen Vertreter aus jeder Domäne, als Mitglied in einen domänenübergreifenden Governance-Council zu entsenden, der hier mit der zentralen IT- und Governanceorganisation zusammenarbeitet und die zwingend notwendige Kommunikation unter- und miteinander unterstützt.[12]

Um dies in der Organisation zu verankern, sind neue Rollen hilfreich: »Domain Data Product Owner« und »Data Product Developer«, welche die Verantwortung für ein Datenprodukt tragen. Dabei geht es darum, Daten als echtes Produkt zu sehen, wie z. B. Softwareartefakte zuvor. Dies führt häufig zu einem Mehraufwand in den jeweiligen Teams, welches weitere Fragen aufwirft. Zum Beispiel, wer für die dafür entstehenden Kosten aufkommt.[13]

3.2 Einführung eines unternehmensweiten Datenkataloges

In der heutigen Welt sind Daten eines der wertvollsten Assets eines jeden Unternehmens. Das Problem, mit dem die meisten Unternehmen konfrontiert sind, besteht jedoch darin, dass die Daten über unzählige Quellen verstreut sind, was ihre effektive Verwaltung und Nutzung erschwert. An dieser Stelle kommt ein Datenkatalog ins Spiel, ein Verzeichnis mit Daten über Daten. Ein Datenkatalog ist ein zentrales Repository, welches die Datenbestände eines Unternehmens organisiert und verwaltet und es den Benutzenden erleichtert, die benötigten Daten(-produkte) zu finden und zu nutzen. Das »Schaufenster« eines Datenkataloges bilden in der Regel Datenprodukte, die für die Attraktivitätskomponente im Kontext der Datentransparenz stehen und für die Datenteilung geschnitten werden, was letztendlich die Kernaufgabe eines Datenkataloges widerspiegelt.

Die Einführung eines unternehmensweit anwendbaren Kataloges bedingt jedoch im Kern fünf Herausforderungen, die früh adressiert werden müssen.

1. Verstehen der Vorteile eines Datenkatalogs

Der erste Schritt bei der Einführung eines Datenkatalogs besteht darin, die Benefits einzuordnen, die dieser für die Organisation bietet. Ein Datenkatalog erleichtert das Auffinden und die Nutzung von Daten, reduziert Duplikate, kann die (Meta-)Datenqualität verbessern und stellt eine hinreichende Governance sicher. Außerdem verbessert er die Zusammenarbeit, die Datenkompetenz und den Wissensaustausch zwischen den Benutzenden. Die Mehrwerte müssen jedoch auch in der Organisation vermittelt werden, da, wie bei jeder Innovation, es oft kulturelle Hürden gibt und ein Datenkatalog nicht als Selbstzweck ausgerollt wird.

12 Bitkom. 2020. Data Mesh – Datenpotenziale finden und nutzen.
↗ https://www.bitkom.org/Bitkom/Publikationen/Data-Mesh-Datenpotenziale-finden-und-nutzen
13 id.

2. Definition der Anforderungen an den Datenkatalog

Umso konkreter die Anforderungen, umso leichter lassen sich diese adressieren! Es ist insbesondere festzulegen, wie die Daten klassifiziert werden sollen, welche Standards angewendet werden und welche (Daten-)Rollen ihren Platz in der Data Governance Organisation finden.

3. Wählen des richtigen Tools

Es gibt viele Datenkataloglösungen auf dem Markt, und es muss die richtige Lösung für die Organisation ausgewählt werden (oder selbst entwickelt werden). Wichtig sind die Funktionen (bspw. eine automatisierte Dokumentation der technischen Metadaten einer neuen Datenquelle), die Benutzerfreundlichkeit, die Skalierbarkeit, die Sicherheit, die Integration und die Kosten.

4. Stakeholdermanagement

Die Einführung eines Datenkatalogs ist keine technische Aufgabe, sondern erfordert die Einbeziehung von Beteiligten aus verschiedenen Bereichen des Unternehmens. Geschäftsanwender, insbesondere Data Owner, sollten in den Prozess einbezogen werden. Das Verstehen der Bedürfnisse, Bedenken und Erwartungen der Kunden ist hier essenziell.

5. Durchführung der Implementierung

Wenn die Roadmap steht, ist es an der Zeit, die Implementierung anzugehen. Die Datenkataloglösung wird bereitgestellt, die ersten Datenquellen werden angebunden, die ersten Datenmodelle werden aufgebaut. Hier ist die Schulung der Nutzenden, vor allem der »Poweruser«, die wiederum andere interessierte Kolleginnen und Kollegen schulen können (Multiplikatoransatz), essenziell. Auch hier ist es relevant, die Mehrwerte des Kataloges zu kommunizieren.

Sicherlich ergibt es zudem Sinn, ein Proof of Concept bzw. Use Case in einem Geschäftsfeld zu entwickeln, um hier eine Success Story bzw. Blaupause zu kreieren. So können auch Geschäftsfelder angesprochen werden, die zurzeit noch Bedenken gegen die Einführung / Verwendung eines Datenkataloges haben.

3.3 Wechselwirkung zwischen Datenkatalog und Datenprodukt

Datenproduktion: Jetzt geht es erst richtig los!

Die Einführung eines Datenkatalogs ist keine einmalige Aufgabe, sondern ein kontinuierlicher Prozess. Die Nutzung, das Feedback und die Leistung des Datenkatalogs sollten überwacht werden. Wichtig ist es, verbesserungswürdige Funktionen zu überprüfen und entsprechende Änderungen zu implementieren. Es muss sichergestellt werden, dass sich der Datenkatalog mit den sich ändernden Anforderungen des Unternehmens weiterentwickelt. Hilfreich ist die »Inventarisierung« von Datenbeständen und Datenprodukten durch den Katalog auch insofern, als dass möglicherweise intern verfügbare Datenbestände oder -produkte nicht extern beschafft werden müssen. Hierbei ist zwischen **Datenbeständen** und **Datenprodukten** zu unterscheiden. Das Primärziel des Datenkatalogs liegt darin, Datenbestände auffindbar und verfügbar zu machen. Dies kann zu einer schnelleren und genaueren Entscheidungsfindung, einer verbesserten abteilungsübergreifenden Zusammenarbeit und einem besseren Verständnis der Datenbestände des Unternehmens führen. Wir halten fest: Ein Datenkatalog ergibt auch ohne bereits entwickelte Datenprodukte Sinn. Ein singuläres Befüllen einfacher Datenquellarchitekturen (ganze Datenbanken, Datenschemata, einzelne Tabellen) wird allerdings wenig Nutzen bei möglichen Kundinnen und Kunden des Datenkatalogs erzeugen, da diese Daten schwer zu interpretieren sind, und die Wiederverwendungsmöglichkeiten sehr gering erscheinen.

An dieser Stelle kann die Verbindung zwischen Datenbeständen und Datenprodukten gezeigt werden. Die Auffindbarkeit und Verfügbarkeit von Datenbeständen unterstützt die Entwicklung von Datenprodukten. Dazu werden Datenprodukte im Datenkatalog registriert. Durch die Registrierung von Datenprodukten in einen Datenkatalog können Unternehmen den Benutzenden einen umfassenderen Überblick über die verfügbaren Datenprodukte und deren (aktuelle) Verwendung bieten.

Darüber hinaus können Datenprodukte innerhalb eines Datenkatalogs verwendet werden, um den Wert der Datenbestände des Unternehmens gegenüber Interessengruppen wie Kundinnen und Kunden, Partnerinnen und Partner und Investorinnen und Investoren zu demonstrieren. Sie können die Erkenntnisse, Trends und Vorhersagen aufzeigen, die sich aus den Datenbeständen ableiten lassen, und demonstrieren, wie Daten zur Verbesserung der Geschäftsergebnisse genutzt werden können.

Es lässt sich zusammenfassen, dass Datenprodukte innerhalb eines Datenkatalogs wichtig sind, weil sie dem Unternehmen und seinen Nutzenden einen Mehrwert bieten und helfen, den Wert der Datenbestände des Unternehmens zu demonstrieren. Sie ermöglichen eine schnellere und genauere Entscheidungsfindung, verbessern die Zusammenarbeit und zeigen die Erkenntnisse und Vorhersagen, die aus den Daten abgeleitet werden können.

3.4 Datenprodukte im Datenkatalog

Mit der Bereitstellung einer Datenkatalog-Lösung beginnt der eigentliche Kraftakt erst, diese mit den relevanten Informationen über die Datenprodukte zu befüllen. Diese Arbeit ist keine einmalige Aufgabe und kann auch nicht von einer einzelnen Person durchgeführt werden. Die gesamte Data Community ist daran beteiligt. Der Datenkatalog ist ein lebendes Verzeichnis, welches den aktuellen Stand und Qualität der Datenprodukte reflektiert. Deshalb müssen die Inhalte regelmäßig überprüft und aktualisiert werden, um sicherzustellen, dass die Inhalte korrekt, komplett und relevant sind. Die klare Empfehlung ist der Aufsatz eines Data Governance Frameworks, welches dazu dient, Prozesse, Verantwortlichkeiten und Standards zu definieren und zu etablieren.
Mit der verteilten Verantwortung durch ein Federated Governance-Modell, ist die domänenübergreifende Anwendung von globalen Standards und Richtlinien notwendig. Für den Datenkatalog ist ein Metadatenstandard zu definieren, welcher zugrunde legt, welche Informationen zu einem Datenprodukt in dem Datenkatalog gepflegt werden. Die Spezifikation trifft semantische Regelungen für die Kommunikation und Umsetzung der Governance. Ein Metadatenstandard stellt sicher, dass Datenprodukte alle nach dem gleichen Muster dokumentiert werden und auch für die Datenkonsumentinnen und -konsumenten in einem geeigneten Format zur Verfügung stehen – »Fit for Purpose«.

Nachfolgend werden relevante Ausprägungen an einen solchen Metadatenstandard betrachtet. Die Ausprägungen sind von der gewählten Datenkataloglösung agnostisch und orientieren sich an dem »Data Catalog Vocabulary« (DCAT),[14] veröffentlicht vom World Wide Web Consortium (W3C). DCAT ist ein Vokabular, welches ein Schema definiert, und dazu dient, die Interoperabilität zwischen Datenkatalogen zu gewährleisten.

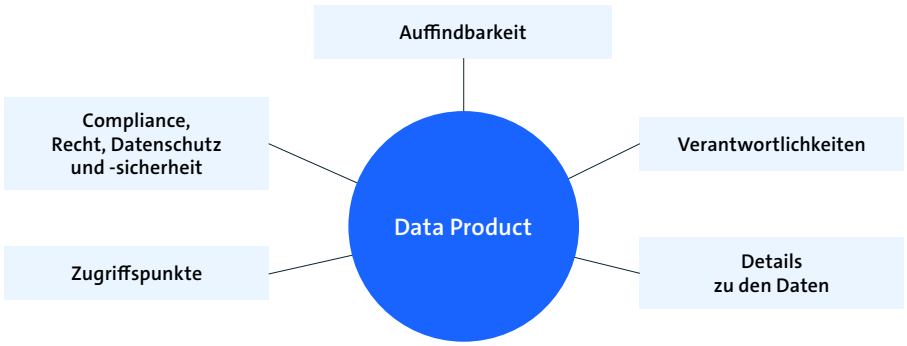


Abbildung 2 – Ausprägungen eines Metadatenstandards für Datenprodukte

Auffindbarkeit

Nachfolgende Attribute dienen dazu, ein Datenprodukt für Konsumentinnen und Konsumenten verständlich darzustellen. Es ist dabei darauf zu achten, dass man keine Fachsprache seiner Domäne verwendet, sondern eine Sprache wählt, die auch Ausstehenden ein leichtes Verständnis gewährleistet.

14 W3. 2020. Data Catalog Vocabulary (DCAT) – Version 2. ↗https://www.w3.org/TR/vocab-dcat-2

Besonders relevant ist, dass ein Datenprodukt mit weiteren Kontextinformationen verknüpft wird. Das ermöglicht, dass die Datenprodukte schneller gefunden werden können. Daten haben in der Regel die Eigenschaft, dass sie sich nicht selbst erklären. Die Konsumentinnen und Konsumenten können die Datenprodukte über Ihnen bekannte Strukturen und Begriffe auffinden, etwa über die Verknüpfung mit einem Geschäftsobjekt, welches ermöglicht, Daten zu einem Objekt aus der realen Welt zu finden. Dabei können Verknüpfungen auf bestehende Elemente aus der Enterprise Architecture verwendet werden, wie Geschäfts-Objekte, Prozesse, Domänenmodell etc.

- Identifier
- Titel
- Beschreibung
- Tags / Keyword
- Kontextinformationen
  - Geschäftsobjekte[15]
  - Prozesse
  - Domänen
  - ...

Verantwortlichkeiten

Die Data Governance hat als ein Ziel die klare Zuweisung von Verantwortlichkeiten im Datenmanagement. Die Verantwortlichkeiten können im Datenkatalog an dem Datenprodukt dokumentiert werden. Das ermöglicht den Konsumentinnen und Konsumenten direkten Kontakt zu den Verantwortlichen aufzunehmen und spart die lange Suche nach den Kontakten. Dies ist enorm wichtig durch die Dezentralisierung, die der Data Mesh[16] Ansatz mit sich bringt.

Zu den spezifischen Rollen existieren vielfältige Ausprägungen in der Literatur und in den Unternehmen. Es ist zu empfehlen, keine reine technische oder fachliche Zuordnung zu setzen, sondern auf eine geteilte Verantwortung. Nachfolgend sind zwei Rollen aufgeführt, die die wesentlichen Teile des Rollenkonzeptes abdecken. Bei Bedarf können weitere Rollen gepflegt werden.

- Data Owner
- Data Steward
- Weitere Rollen

Details zu den Daten

Die Details zu den Daten sind für die mehr technisch avisierten Nutzenden, die später die Daten verarbeiten werden. Es soll aufgezeigt werden, welche Daten und in welcher Struktur das Datenprodukt zur Verfügung stellt. Dazu eignet sich Informationen über das Datenschema, was eine formale Beschreibung der Struktur von Daten ist, bereitzustellen. Neben einer formalen Beschreibung hilft auch ein beschreibendes Beispiel, z. B. wenige Zeilen, die das Datenprodukt bereitstellt. Um Transparenz und Vertrauen für ein Datenprodukt herzustellen, ist es ratsam, ein Qualitätssiegel bereitzustellen. Dazu kann auf Kennzahlen zur Messung der Datenqualität zurückgegriffen werden. Details zu dem Thema Datenqualität kann dem ↗ Kapitel 4 entnommen werden.

15 Wikipedia. 2023. Geschäftsobjekt, ↗ https://de.wikipedia.org/wiki/Gesch%C3%A4ftsobjekt
16 Dehghani, Z. 2022. Data Mesh. O'Reilly, ↗ https://www.oreilly.com/library/view/data-mesh/9781492092384/

- Quelle
- Datenschema
- Beschreibendes Beispiel
- Datenqualität
- Datenkategorie

Zugriffspunkte

Der Zugriffspunkt oder auch Distribution ist eine physische Repräsentanz des Datenproduktes in einem spezifischen Format. Ein Datenprodukt kann mehrere Zugriffspunkte haben.

- Zugangs-URL
- Format
- Verfügbarkeit
- Bereitstellungsintervall

Compliance, Recht, Datenschutz- und Sicherheit

Ein Unternehmen muss wissen, welche sensible Daten verarbeitet werden und auch wo diese verortet sind. Durch eine standardisierte Klassifizierung der Daten und der Dokumentation an einer zentralen Stelle kann trotz des dezentralen Ansatzes die Transparenz sichergestellt werden. Diese Transparenz ist ein entscheidender Baustein, um die Sicherheit und Compliance einzuhalten und die Grundlage, um Vertrauen zu schaffen.

Durch die Klassifizierung werden auch die Rechte und Rahmenbedingungen für die Konsumentinnen und Konsumenten zur Nutzung der Daten festgelegt. Datenprodukte sind mit einer Lizenz auszuweisen. Neben der Lizenz können Angaben über Nutzungsbestimmungen und zu Regelwerken / Richtlinien gemacht werden.

- Lizenz
- Nutzungsbedingungen
- Löschfristen
- Richtlinien
- Schutzbedarf
- Personenbezogene Daten

Mit dem hier vorgeschlagenen Metadatenstandard für Datenprodukte in einem Datenkatalog wird eine übergreifende Struktur sichergestellt, welche alle Domänen folgen. Die hier genannten Ausprägungen stellen sicher, dass die Anforderungen an ein Datenprodukt erfüllt werden, siehe ↗ Kapitel 1.

3.5 Data Marketplace: mit Daten handeln

Was sind Datenmarktplätze und wie funktionieren sie?

Ein Datenmarktplatz ist eine Plattform, auf der Nutzende verschiedene Arten von Datensätzen und Datenströmen aus verschiedenen Quellen kaufen oder verkaufen können. Diese Plattformen ermöglichen einen Self-Service-Datenzugriff und gewährleisten gleichzeitig Sicherheit, Konsistenz und hohe Datenqualität für die Parteien.

Der Vorteil von Datenmarktplätzen in Unternehmen besteht in der Wertschöpfungskette für Käuferinnen und Käufer sowie Verkäuferinnen und Verkäufer. Daten können genutzt werden, um den Datenzugriff und die Datenanalyse zu verbessern bzw. sicherer zu gestalten.

Ein Datenmarktplatz kann innerhalb eines Unternehmens verwendet werden, um Daten-Konsumenten ein intuitives, sicheres, zentralisiertes und standardisiertes Daten-Einkaufserlebnis zu bieten. Er bringt Daten den Datenanalystinnen und -analysten und Wissenschaftlerinnen und Wissenschaftlern näher, indem es die zugrunde liegenden Metadaten nutzt.

In Verbindung mit einem Data Catalog-Dienst[17] kann ein Datenmarktplatz umfangreiche Suchfunktionen bieten, mit denen Benutzer nach Schlüsselwörtern, Geschäfts-Begriffen und natürlichen Sprachen suchen können. Azure Purview zum Beispiel unterstützt die Datenermittlung, einschließlich Glossaren und Klassifizierungen, sodass Datennutzende Daten leicht finden können.

Eine externe Nutzung kann über Datenmarktplätze für Einzelpersonen und Organisationen sinnvoll sein, um Daten anzureichern und gemeinsam zu analysieren.

Um die Datenermittlung und den Datenzugriff zu verbessern, können – wo technisch möglich – Datenmarktplätze in Data-Catalogue-Dienste integriert werden, um sowohl interne als auch externe Datenquellen einfach einzubinden.

An dieser Stelle kann das Konzept eines Datenraums[18] (»Dataspace«) als Synonym verwendet werden – ein Datenraum hat typischerweise das Ziel, einen Data Management und Data Governance Layer zwischen Organisationen bereitzustellen, über welchen selbstbestimmt Daten geteilt (gehandelt) werden können.

Es gibt aber auch Risiken, die mit dem Verkauf von Daten sowohl innerhalb als auch außerhalb eines Unternehmens verbunden sind. Ein Risiko besteht darin, dass die verkauften Daten möglicherweise nicht korrekt oder von vereinbarter Qualität sind.

17 z. B. Azure Purview, AWS Glue, atlan, Google Data Catalog, Oracle Data Catalog, IBM Watson Knowledge Catalog, SAP Data Intelligence Catalog, Cloudera Data Catalog.

18 Bitkom. 2022. Datenräume und Datenökosysteme: Erste Einordnung und aktueller Stand, ↗ https://www.bitkom.org/Bitkom/Publikationen/Datenraeume-Datenoekosysteme-erste-Einordnung-aktueller-Stand

Darüber hinaus gibt es rechtliche Risiken, die mit dem Verkauf von Daten verbunden sind. Unternehmen müssen Datenschutzbestimmungen einhalten, wie z. B. die Datenschutz-Grundverordnung (DS-GVO) der EU oder den California Consumer Privacy Act (CCPA), die Regeln, wie Daten gesammelt, verwaltet, weitergegeben und verkauft werden. Die Nichteinhaltung dieser Vorschriften kann insbesondere zu hohen Bußgeldern und Marken-Schäden führen.

Ein weiteres Risiko besteht darin, dass es häufig ein Misstrauen gegenüber Unternehmen gibt, die ihre Daten verkaufen. Dies kann zu negativer Publicity und Reputationsschäden führen.

Um diese Risiken zu vermeiden, ist das Management von Datenqualität- und Sicherheit, bei der Veräußerung von Daten, von entscheidender Bedeutung. Die Datenqualität kann durch die Implementierung von Datenqualitätsmetriken sichergestellt werden, um die Qualität von Datenprodukten zu bewerten und zu verbessern. Diese Metriken können Vollständigkeit, Eindeutigkeit, Konsistenz, Gültigkeit, Genauigkeit und Verknüpfung umfassen.

Datensicherheit kann durch die Implementierung durchdachter Mechanismen und Systeme zur Gewährleistung des Datenschutzes erreicht werden. Dies sollte Verschlüsselung, Zugriffskontrollen und regelmäßige Sicherheitsüberprüfungen umfassen.

Ein Datenmarktplatz ist in der Regel eine dünne Orchestrierungsschicht mit einem ansprechenden Erscheinungsbild, die gutes Benutzererlebnisse bietet. Datenmarktplätze verwenden zugrunde liegende Metadaten-Repositories, bei denen es sich um eine Mischung aus selbst entwickelten Metadaten-Speichern und zum Beispiel Diensten wie Azure Purview, Colibra, SAP-Datasphere oder andere handelt.

3.6 Automatisierungsansätze für Data Governance und Datenqualität

Mit einer steigenden Anzahl an Datenprodukten, Objekten im Datenkatalog und auch Beteiligten sind Fragestellungen rund um den skalierbaren, kosteneffizienten und zuverlässigen Betrieb an einem bestimmten Punkt unabwendbar. Folglich rückt das Thema »Automatisierung« der zugrunde liegenden Abläufe in den Fokus. Aus dem Bereich des Machine Learning, oder auch der Softwareentwicklung sind Frameworks wie DataOps, MLOps oder DevOps mit zugrundeliegenden »CI / CD Pipelines« nicht mehr wegzudenken, und sind wie bereits im vorherigen ↗Leitfaden Data Mesh auch integraler Bestandteil einer nachhaltigen Data Analytics.

Im Kontext Data Governance oder auch Datenqualität erscheint es als logische Konsequenz, diese Punkte auch weitestgehend in ein bestehendes DataOps Framework einzubinden. Im Falle von Datenqualität kann das schon mit einer automatisierten Überwachung der neu ins Modell einfließenden Daten zur Vermeidung nachfolgenden Modell Problemen beginnen.

Unter dem Stichwort »Data Drift Monitoring« wird man in gängigen MLOps Frameworks hierzu fündig. Doch auch die automatisierte Berechnung von Data Quality Scores und Anzeige im Datenkatalog kann hier ein Schritt sein, um den Anwendern direkt einen Blick auf die Qualität eines Datensatzes zu geben.

Darauf aufsetzend können auch automatisierte Abläufe bei der Verwendung und Erstellung von Datenkatalogen vereinfachen. Hier sind zum Beispiel das automatisierte Verschlagworten (Tagging) von Einträgen zu nennen. Die Eintragung von Datenquellen ist sehr arbeitsintensiv, sodass sich in diesem Bereich auch der Einsatz von KI anbietet, um passende Tags vorzuschlagen. Allerdings empfiehlt es sich, diese Vorschläge im Rahmen eines Workflows durch einen Data Steward oder entsprechende Fachabteilungen zu validieren. Ein ähnliches Vorgehen sollte auch für die Klassifikation der Datensätze im Hinblick auf z. B. DS-GVO, Lizenzen oder Nutzungsrechte erfolgen.

Auch bei der konsistenten Einbindung & Verwendung von role-based access control (RBAC) auf alle Ressourcen kann durch Automatisierung die schnelle Verwendbarkeit eines Datenkataloges verbessert werden: Warum nicht gleich die Gruppenzuordnungen aus dem unternehmensweiten Mitarbeiterverzeichnis nutzen, um passende Datenquellen für einen Nutzer vorzuschlagen, oder auch die Berechtigungen für bestimmte Datenprodukte entsprechend einrichten? Da es oftmals um sensitive Daten geht, kann es je nachdem auch hier sinnvoll sein, den Data Steward in den Workflow einzubinden.

Im Hinblick auf Sicherheit ist auf jeden Fall das Monitoring von unautorisierten Zugriffen zu nennen – auch wenn das nicht direkt mit Data Governance im engeren Sinne zu tun hat.

Je nach Unternehmensphilosophie kann es auch sinnvoll sein, die Zuordnung von Kosten, z. B. Cloud-Computing und Storage auf einzelne Datenprodukte automatisiert zu erfassen.

Ähnlich wie bei Modellen, Reports oder auch Dashboards gilt auch bei Einträgen im Datenkatalog bzw. Datenprodukten: Es wird Revisionen, Updates und Verbesserungen geben – daher am besten direkt so aufsetzen, dass dies leicht, effizient, auditierbar und vor allem ohne viel manuellen Aufwand – also automatisiert – geschieht. Das spart Kosten und erleichtert eine spätere Skalierung enorm.

4 Datenqualität steigern und absichern

4 Datenqualität steigern und absichern

4.1 Qualität: Aspekte, Merkmale, Attribute

Angetrieben von dem Glauben, dass mehr Daten mehr Nutzen bringen, konzentrieren sich Unternehmen auf das Sammeln von Daten, ohne sich vorher Gedanken darüber zu machen, was sie mit den Daten erreichen wollen und wie. Unternehmen neigen dazu, zu glauben, dass sie mit einem Mangel an Daten zu kämpfen haben, obwohl in Wirklichkeit die meisten von ihnen mehr als genug Daten haben, um aufschlussreiche Entscheidungen zu treffen. Das eigentliche Problem ist nicht die Menge, sondern die Qualität der Daten. Nach dem Prinzip »Garbage-in-Garbage-out« sind selbst die größten Datensätze nutzlos und verursachen nur Kosten, wenn sie von schlechter Qualität sind. Aber was ist Datenqualität und wie kann man sie managen?

Datenqualität wird im Allgemeinen als »Gebrauchstauglichkeit« (eng. **Fitness for use**) definiert. Diese allgemeine Definition impliziert, dass die Qualität eines Datenprodukts von den Bedürfnissen der Nutzenden und dem Nutzungskontext abhängt.

In der Praxis ist die Datenqualität ein komplexes Konzept, das aus einer potenziell großen Anzahl von oft miteinander verbundenen Qualitätsaspekten besteht, die auch als Qualitätsmerkmale oder Qualitätsattribute bezeichnet werden. ISO / IEC 25012[19] beispielsweise definiert Datenqualität als den »Grad, in dem die Merkmale von Daten die angegebenen und impliziten Bedürfnisse bei der Verwendung unter bestimmten Bedingungen erfüllen«.

Obwohl wir die Qualität von Daten immer aus der Perspektive ihrer spezifischen Verwendung betrachten sollten, können sich einzelne Qualitätsaspekte hinsichtlich ihrer Kontextspezifität unterscheiden. Einige wenige Qualitätsmerkmale werden gewöhnlich als »allgemeingültig« betrachtet, unabhängig von einem bestimmten Verwendungszweck- und Kontext der Daten. Solche Qualitätsaspekte werden als **inhärent** oder **intrinsisch** bezeichnet und geben an, wie gut Daten die realen Fakten repräsentieren, die sie darstellen sollen. Zu den wichtigsten Beispielen für inhärente Datenqualitätsaspekte gehören Datenkorrektheit, Konsistenz und Vollständigkeit. **Kontextuelle** Qualitätsaspekte hingegen sind nur für einen bestimmten Zweck und Kontext der Datennutzung relevant. Dazu gehören beispielsweise Verteilungseigenschaften der Daten – wie die Ausgewogenheit der Werte – oder strukturelle Eigenschaften – wie Formatierung oder Skalierung, die für bestimmte Analysetechniken- und Werkzeuge erforderlich sind.

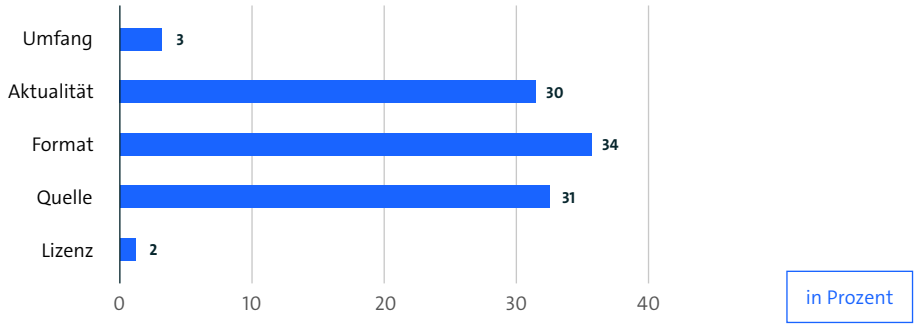
Aus der Perspektive des **Datenlebenszyklus** ist die Unterscheidung zwischen kontextunabhängiger und kontextabhängiger Qualität relevant. Kontextunabhängige Qualität bezieht sich auf Eigenschaften von Daten, die nicht vom Kontext der Datenanwendung abhängen.

19 ISO/IEC. 2008. ISO/IEC 25012:2008. Software engineering — Software product Quality Requirements and Evaluation (SQuaRE) — Data quality model. https://www.iso.org/standard/35736.html

Das bedeutet, dass diese Qualitätseigenschaften zum Zeitpunkt der Datenerfassung sichergestellt werden können – und in der Regel auch sollten – und »zentral« statt »lokal« in jeder Datenanwendung gepflegt werden.

Trotz der Bedeutung der Datenqualität bleibt ein Großteil der Arbeit individuell, handgemacht und qualitativ. In der Literatur werden Konzepte wie die »3 Vs« von Volumen, Geschwindigkeit und Vielfalt verwendet,[20] und es kommen weitere Vs hinzu, wie Variabilität und Wert.[21] Aus operativer Sicht, d. h. aus Sicht einer Analyseanwendung, sind die Vs jedoch konzeptionell und qualitativ geblieben. Die Vs können für eine erste Bewertung nützlich sein, vielleicht für einen Pre-Test, eine erste Datenauswahl in Form einer Triage. Um jedoch die Ergebnisse in Bezug auf die Leistung zu bewerten, die Wahrscheinlichkeit von Effekten (x verbessert y), die Größe von Effekten (x verbessert y um ein Vielfaches) und die Signifikanz (Verbesserungen sind real und nicht zufällig) zu schätzen, enthalten die Vs zu wenig Informationen. Nehmen wir zum Beispiel Verkehrsdaten für eine Routing-App oder Parkdaten für eine Park-App: Wie frisch sind die Daten? Wie häufig werden sie aktualisiert? Oder betrachten Sie Zeitreihendaten: Wie lang und granular sind sie? Wie lang ist der gesamte Beobachtungszeitraum (10 Jahre im Gegensatz zu 1 Jahr), und wie dicht sind die Daten für jeden Zeitraum (Daten eines Jahres in monatlichen, täglichen, stündlichen Intervallen?). Eine erste, explorative Umfrage in Schlueter Langdon & Sikora bestätigt die Möglichkeit der Qualitätsbewertung: Während Qualität gegenüber Quantität bevorzugt wird (Wie gebe ich den nächsten US $1 aus? Qualität: 82 Prozent, Quantität: 18 Prozent; n = 65), zeichnet sich kein eindeutiger Qualitätsindikator ab (Volumen: 3 Prozent, Frische: 30 Prozent, Format: 34 Prozent, Quelle: 31 Prozent, Lizenzart: 2 Prozent, n = 64).[22]

Wie wird Datenqualität gemessen?



Schlueter Langdon, C., and R. Sikora. (2020)

Abbildung 3 – Möglichkeit der Qualitätsbewertung

20 McAffee, A., and E. Brynjolfsson. 2012. Big data the management revolution. Harvard Bus. Rev. 90(10): 60–68
21 z. B. Yin, S., and O. Kaynak. 2015. Big data for modern industry: challenges and trends. Proc. IEEE 103(2): 143–146
22 Schlueter Langdon, C., and R. Sikora. 2020. Creating a Data Factory for Data Products. In: Lang, K. R., J. J. Xu et al. (eds). Smart Business: Technology and Data Enabled Innovative Business Models and Practices. Springer Nature, Switzerland: 43-55

4.2 Qualitätsmodell

Die Identifizierung relevanter Qualitätsmerkmale und die Unterscheidung zwischen inhärenten und kontextabhängigen Merkmalen ist nur ein erstes Element der Spezifikation und des Managements der Datenqualität. Für ein effektives und effizientes Management der Datenqualität ist es notwendig, die Qualität auf eine überprüfbare Weise zu spezifizieren. Das Datenqualitätsmodell bietet einen verständlichen und systematischen Weg, die Qualität von Datenprodukten konsistent zu definieren, zu kommunizieren, zu bewerten und zu kontrollieren. Abbildung 4 fasst die grundlegenden Elemente eines Datenqualitätsmodells und ihre Beziehungen zueinander zusammen.

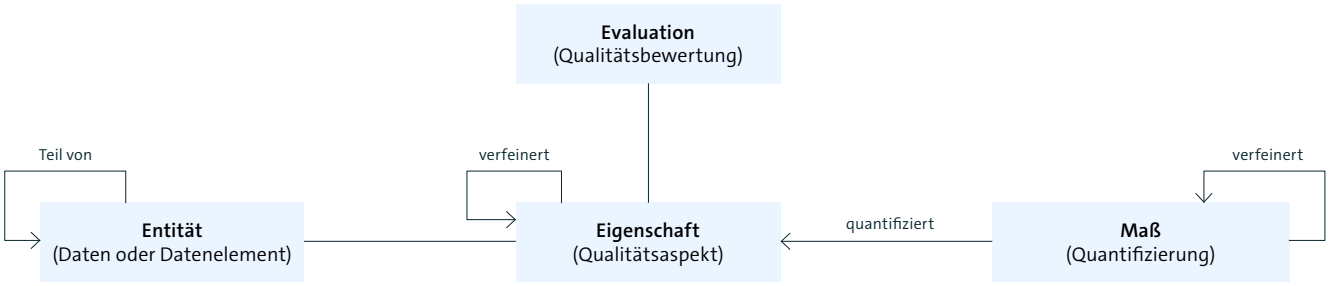


Abbildung 4 – Grundbestandteile eines Datenqualitätsmodells

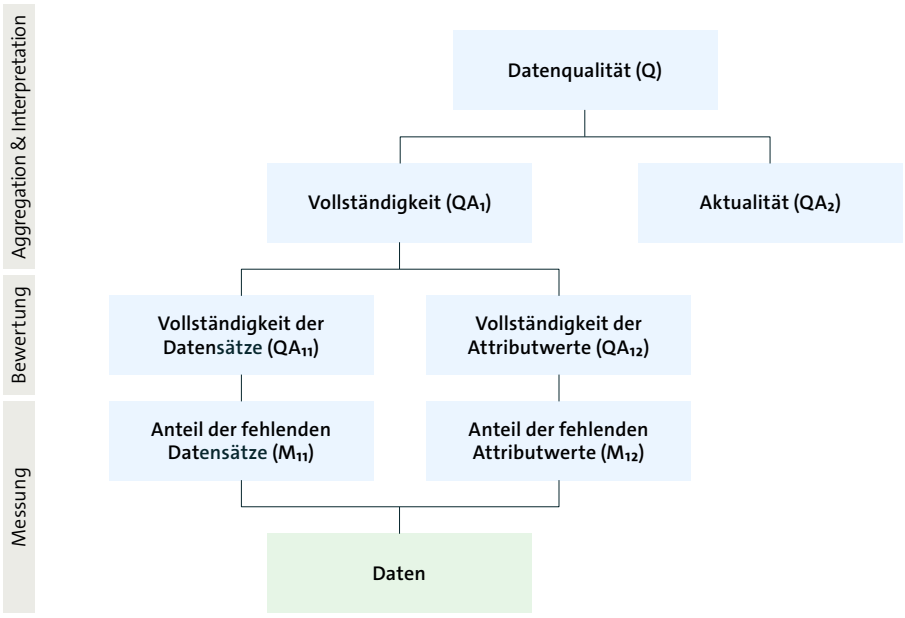


Abbildung 5 – Beispiel für ein einfaches Datenqualitätsmodell

Das zentrale Element eines Qualitätsmodells ist die Eigenschaft einer Entität, d. h. ein Attribut, das die Entität charakterisiert und mit der Qualität der Entität zusammenhängt oder, mit anderen Worten, einen Qualitätsaspekt einer Entität darstellt.

Entität bezieht sich auf ein Datenprodukt, einschließlich Daten und deren Elemente, wie z. B. Datensätze oder Attribute. Um die Datenqualität eindeutig zu definieren und zu bewerten, müssen abstrakte Eigenschaften und Entitäten durch die Relationen »refine« und »part-of« in spezifischere, messbare Eigenschaften und Entitäten heruntergebrochen werden. So kann beispielsweise (Abbildung 5) die »Vollständigkeit« eines Datenprodukts in die Vollständigkeit von Datensätzen und die Vollständigkeit von Attributwerten unterteilt werden.

Das **Maß** definiert, wie eine bestimmte Eigenschaft (einer bestimmten Entität) in einem bestimmten Kontext quantifiziert wird. Zum Beispiel kann die »Vollständigkeit der Attributwerte« einer Kundendatenbank als Anteil der fehlenden Geburtsdateneinträge in den Kundendatensätzen gemessen werden.

Die **Evaluation** umfasst vier grundlegende Elemente: Messung, Bewertung, Aggregation und Interpretation.

Die **Messung** besteht aus der Erfassung von Messdaten für die Eigenschaften von Datenelementen, die auf der untersten Ebene der Hierarchie des Qualitätsmodells gemäß den im Qualitätsmodell definierten Maßen angegeben sind.

Die **Bewertung** umfasst die Beurteilung der Erfüllung der vordefinierten Qualitätspräferenzen (Anforderungen), die mit bestimmten Eigenschaften verbunden sind. Im Beispiel in Abbildung 5 wird die Präferenz bezüglich »Fehlender Attributwerte« (M_{12}) durch eine Bewertungsfunktion im Qualitätsaspekt (QA_{12}) »Vollständigkeit der Attributwerte« definiert. Im einfachsten Fall kann die Bewertungsfunktion durch einen Schwellenwert definiert werden, der den maximal akzeptablen Anteil fehlender Attributwerte darstellt, oberhalb dessen QA_{12} nicht akzeptabel ist, d. h. die Daten haben die schlechteste Qualität in Bezug auf dieses spezifische Merkmal.

Die **Aggregation** umfasst die Synthese der Bewertungen, die auf der Grundlage von Messungen und Bewertungen in der gesamten Qualitätsmodellhierarchie von unten nach oben zu einer Gesamtqualitätsbewertung vorgenommen werden. Das Qualitätsmodell in Abbildung 5 stellt beispielsweise einen sogenannten kompensatorischen Ansatz dar, bei dem schlecht abschneidende Qualitätsmerkmale bis zu einem gewissen Grad durch gut abschneidende Merkmale kompensiert werden können. In der Praxis würden wir jedoch sogenannte »Veto«-Schwellenwerte (t_{11} und t_{12}) für die einzelnen Qualitätsmerkmale festlegen, die bei Überschreitung eines dieser Werte zu einer nicht vertretbaren Gesamtqualität führen würden (»Veto einlegen«).

Schließlich besteht die **Interpretation** darin, die potenziell abstrakten Qualitätsbewertungen in für menschliche Entscheidungsträger verständliche (intuitive) Bewertungen zu übersetzen. Diskrete Ebenen, wie z. B. das dreistufige Ampelsystem, sind sehr verbreitet, um einen schnellen Überblick über die aktuelle Datenqualität zu erhalten.

Bevor eine Datenqualitätsbewertung durchgeführt werden kann, muss sie zunächst operationalisiert werden. Der Schritt der Messung kann die Definition zusätzlicher

Maße erfordern, um sicherzustellen, dass Messdaten, die für dasselbe Maß bei ähnlichen Datenprodukten erhoben wurden, vergleichbar sind. Der Bewertungsschritt erfordert die Definition von Präferenzfunktionen zur Modellierung der Entscheidungskriterien in Bezug auf die für die Eigenschaften definierten Maße. Im einfachsten Fall kann eine Präferenz als Akzeptanzschwelle (Ziel) modelliert werden, d. h. als ein spezifischer Grenzwert, der bestimmt, welche Werte einer Kennzahl bevorzugt (akzeptabel) sind und welche nicht. Die Aggregation erfordert die Definition des Aggregationsoperators, um die Bewertungen der einzelnen Eigenschaften von Entitäten in der gesamten Qualitätsmodellhierarchie zu einer Gesamtbewertung zusammenzufassen. Dies kann die Festlegung von Präferenzen hinsichtlich der relativen Bedeutung von Qualitätsunteraspekten beinhalten, die für denselben Qualitätsaspekt in der Qualitätsmodellhierarchie definiert wurden, um mögliche Entscheidungskompromisse zwischen den Unteraspekten zu berücksichtigen. Die relative Wichtigkeit von Eigenschaften von Entitäten kann durch numerische Gewichte quantifiziert werden. Die Interpretation schließlich erfordert die Definition eines Interpretationsmodells, das Evaluationsergebnis in eine verständliche Bewertung übersetzt, die von Entscheidungsträgern richtig interpretiert werden kann, um beispielsweise geeignete Verbesserungsmaßnahmen abzuleiten. Die Nutzer des Qualitätsmodells und der Qualitätsbewertungsmethode können (und sollten) vor der Qualitätsbewertung eine Operationalisierung vornehmen, um den Ansatz an ihren spezifischen Kontext anzupassen; sie sollten beispielsweise Präferenzfunktionen und Gewichtungen anpassen, um ihre spezifischen Präferenzen hinsichtlich der relativen Bedeutung einzelner Eigenschaften von Entitäten widerzuspiegeln. Der Wissensbereich der multikriteriellen Entscheidungsanalyse (eng. Multicriterial Decision Analysis, MCDA[23]) bietet eine Reihe von Techniken zur Operationalisierung von Datenqualitätsmodellen.

4.3 Qualitätsmanagement für Datenprodukte

Die außerordentliche Bedeutung von Daten für den unternehmerischen Geschäftserfolg und die wachsende Betrachtung von Daten als Produkte eines Unternehmens erfordert ein passendes Qualitätsmanagement. In Produktionsumgebungen hat sich hierfür das Total Quality Management (TQM)[24] als Ansatz etabliert. Darauf aufbauend ist der »Total Data Quality Management« (TDQM) Prozess entstanden, welcher vier Schritte für das Qualitätsmanagement von Datenprodukten definiert.[25] In jedem Prozessschritt gibt es dabei Möglichkeiten zur Automatisierung durch Tools.

Für eine erfolgreiche Bewirtschaftung von Datenprodukten sowie ein gezieltes Datenqualitätsmanagement ist eine ganzheitliche Betrachtung des Datenqualitätsmanagementprozesses notwendig. Eine einzelne Betrachtung von Prozessschritten (z. B. der Messung) reicht nicht aus, die Anforderungen verschiedener Akteure zu erfüllen. Darüber hinaus ergibt es Sinn, den Datenqualitätsprozess sowie dessen Realisierung regelmäßig zu überprüfen und im Rahmen des Geschäftsprozessmanagements adäquat zu managen.[26]

23 Wikipedia. 2023. Multicriterial Decision Analysis, ↗ https://en.wikipedia.org/wiki/Multiple-criteria_decision_analysis
24 Wikipedia. 2023. Total Quality Management, ↗ https://en.wikipedia.org/wiki/Total_quality_management
25 Wang, R. Y. 1998. A product perspective on total data quality management. Communications of the ACM, 41(2), 58–65.
26 Pipino, L. L., Lee, Y. W., & Wang, R. Y. 2002. Data quality assessment. Communications of the ACM, 45(4), 211–218.

Die nachfolgend dargestellte Prozessübersicht fasst die vier Prozessschritte sowie beispielhafte Möglichkeiten zur Automatisierung zusammen.

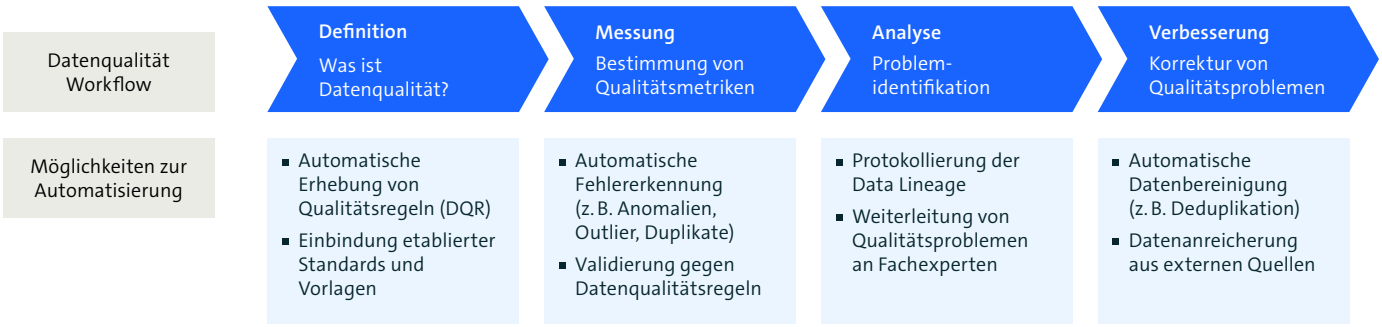


Abbildung 6 – Übersicht für einen Datenqualitätsprozess auf Basis von Wang, R. Y. (1998)[27]

Definition

In diesem Schritt wird ein kontextspezifisches Qualitätsmodell definiert, inkl. Definition der im Anwendungsfall relevanten Qualitätscharakterstiken, Qualitätsmetriken, und Qualitätsevaluation Ansatz.

Messung

Der zweite Schritt befasst sich mit Evaluation der Datenqualität. Hierzu werden Qualitätsmessdaten mit Hilfe geeigneter Messmethoden erhoben. Anschließend werden die einzelnen Qualitätsmerkmale bewertet, aggregiert und interpretiert.

Analyse

Der Analyseschritt umfasst die Identifizierung von Qualitätsdefiziten und die Bewertung ihrer Dringlichkeit. Außerdem erfolgt in diesem Schritt die Identifizierung der Probleme, die einer mangelnden Datenqualität zugrunde liegen (sog. Root-Cause-Analyse[28]).
Qualitätsdefizite werden auf bestimmte Qualitätsmerkmale zurückgeführt, die in einem Qualitätsmodell definiert sind. Die zugehörigen Prozesse und Werkzeuge zur Datenerfassung-, Speicherung- und Pflege werden auf mögliche Ursachen für die festgestellten Qualitätsmängel untersucht.

Verbesserung

Im vierten Schritt werden die identifizierten Datenqualitätsprobleme behoben. Dies kann entweder durch die direkte Modifikation von Daten (d. h. Datenaufbereitung) oder die Anpassung von Prozessen und Tools, die Daten generieren, geschehen. Auch externe Datenquellen können hierbei zur weiteren Verbesserung und Anreicherung des eigenen Datensatzes genutzt werden.

Ein gezieltes und passgenaues Datenqualitätsmanagement ist sehr komplex. Die Komplexität wird insbesondere durch die Subjektivität der Datenqualitätsdefinition und der Vielzahl an Metriken sowie durch heterogene Daten- und Systemlandschaften geprägt. Um diese Komplexität beherrschbar zu machen, sind in der Wissenschaft und Praxis eine Vielzahl an Datenqualitätstools entstanden.

27 Wang, R. Y. 1998. A product perspective on total data quality management. Communications of the ACM, 41(2), 58–65.
28 Wikipedia. 2023. Root Cause Analysis, ↗ https://en.wikipedia.org/wiki/Root_cause_analysis

4.4 Toolgestütztes Qualitätsmanagement

Datenqualitätstools sind lange etabliert und spielen für das effiziente Management von Datenqualität sowie bei der technischen Unterstützung der einzelnen Schritte eine wesentliche Rolle. Häufig unterstützen Tools hierbei das Monitoring von Datenquellen, die Definition von Qualitätsmetriken, oder die Fehleranalyse. Zur verbesserten Automatisierung spielen KI und ML-gestützte Verfahren in den verschiedenen Funktionalitäten von Datenqualitätstools eine immer größere Rolle. Beispielsweise kann die Identifizierung von eindeutigen und nahezu eindeutigen Duplikaten vollständig automatisiert durchgeführt werden.[29]

Etablierte und kommerzielle Datenqualitätslösungen verfolgen im Datenqualitätsmanagement oftmals einen zentralisierten Ansatz. Hierbei ist ein dediziertes Team für das Datenqualitätsmanagement und die Bereinigung der Datensätze verantwortlich. Durch die zunehmende Dezentralisierung der Datenhaltung und des Datenmanagements im Rahmen von »Data Mesh«-Konzepten stoßen etablierte Lösungen jedoch an ihre Grenzen, insbesondere hinsichtlich Skalierbarkeit und Zugänglichkeit. Um in diesen neuen Architekturmodellen erfolgreich zu sein, müssen Datenqualitätstools die Möglichkeit bieten, sich den jeweiligen Anwendungsdomänen (d. h. individuelle Metriken und Datenquellen) anzupassen und gleichzeitig unternehmensweite Standards zu befolgen. Eine Möglichkeit zur Erreichung dieses Ziels ist die Etablierung von unternehmensinternen Datenökosystemen, in denen auf einer technologischen Basis Möglichkeiten zur einfachen Integration von Datenqualitätsapps in Data Sharing Prozesse gegeben werden (siehe Abbildung 7). Auf dieser Basis können Datenprodukte qualitätsgeprüft zur Verfügung gestellt und Datenkonsumenten über die Qualität von Daten informiert werden.[30] Insgesamt sind an dieser Stelle weitere konzeptionelle und technische Arbeiten notwendig, um Datenqualitätstools insgesamt anwendungsfreundlicher, anpassbarer und leichter integrierbar zu gestalten, etwa um diese auch in Datenmarktplätzen zu nutzen.

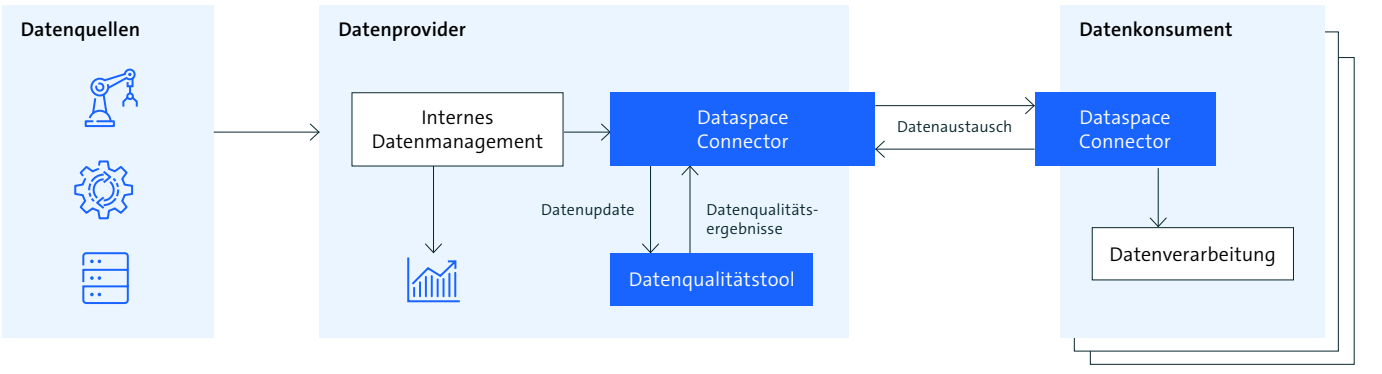


Abbildung 7 – Übersicht Data Sharing Prozesse auf Basis von Altendeitering, M et al. (2022)[31]

29 Altendeitering, M., & Tomczyk, M. 2022. A functional taxonomy of data quality tools: Insights from science and practice. Wirtschaftsinformatik 2022 Proceedings.

30 Altendeitering, M., Pampus, J., Larrinaga, F., Legaristi, J., & Howar, F. 2022. Data sovereignty for AI pipelines: lessons learned from an industrial project at Mondragon corporation. In Proceedings of the 1st International Conference on AI Engineering: Software Engineering for AI (pp. 193–204).

31 Altendeitering, M., Pampus, J., Larrinaga, F., Legaristi, J., & Howar, F. 2022. Data sovereignty for AI pipelines: lessons learned from an industrial project at Mondragon corporation. In Proceedings of the 1st International Conference on AI Engineering: Software Engineering for AI (pp. 193-204).

5 Rechtliche Aspekte

5 Rechtliche Aspekte

Zur rechtlichen Einordnung von Datenprodukten und den damit verbundenen rechtlichen Anforderungen an Art, Inhalt und Qualität der Datenprodukte sowie zur Bestimmung notwendiger Regelungen in datenbezogenen Verträgen ist es zunächst erforderlich, den allgemeinen Rechtsrahmen für Daten darzustellen.

5.1 Kein Dateneigentum

Die Rechtswissenschaft hat sich bereits vor einigen Jahren intensiv mit der Frage auseinandergesetzt, ob es ein Eigentum an Daten gibt. Diskutiert wurden eine analoge Anwendung der Eigentumsregelungen für Sachen, urheberrechtlicher Schutz, Daten als sogenannte Rechtsfrüchte, Zuordnungen über das Datenbankrecht, das Datenschutzrecht oder das Recht zum Schutz von Geschäftsgeheimnissen oder das Konzept einer Datenverfügungsbefugnis.[32] Die Diskussion kam nach der herrschenden Meinung zu dem Ergebnis, dass nach geltendem Recht kein Dateneigentum existiere und die Zuordnung von Daten, ausgehend von einer rein faktisch technischen Zuordnung, über vertragliche Vereinbarungen ausgestaltet werden müsse.

Getrieben durch die technologische Entwicklung, nicht zuletzt im Bereich der künstlichen Intelligenz, haben Daten in den vergangenen Jahren nochmals deutlich an ökonomischer Relevanz gewonnen. Dies und eine zu hohe Konzentration an datenbezogener Marktmacht bei wenigen (Technologie-)Unternehmen haben die EU-Kommission dazu bewogen, regulatorisch einzugreifen und den freien Verkehr von Daten nicht vollständig der vertraglichen Gestaltung zu überlassen. In ihrer europäischen Datenstrategie[33] hat die EU die insoweit geplanten Initiativen skizziert und mit dem Data Governance Act und dem Data Act bereits mit der Umsetzung begonnen. Gleichwohl wird der vertraglichen Gestaltung auch über diese Rechtsakte hinaus weiterhin eine große Bedeutung zukommen, da die Rechtsakte zum einen zwar zwingende Vorgaben für die Vertragsgestaltung enthalten, ein gewisser Gestaltungsspielraum jedoch noch verbleibt, und zum anderen Datenprodukte entstehen werden, die nicht in den Anwendungsbereich der Regelungen fallen (bspw. aus Rohdaten abgeleitete Daten).

32 Siehe hierzu: Grützmacher, CR 2016, 485-495; Dorner, CR 2014, 617-628.

33 Europäische Kommission. 2020. Mitteilung der Kommission an das Europäische Parlament, den Rat, den Euro-päischen Wirtschafts- und Sozialausschuss und den Ausschuss der Regionen – Eine Europäische Datenstrate-gie.COM(2020) 66 final. ↗ https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_de

5.2 Datenschutz

Datenprodukte können personenbezogene sowie nicht-personenbezogene Daten enthalten.

Die datenschutzrechtlichen Regelungen sind auf nicht-personenbezogene Daten nicht anzuwenden, sodass Datenprodukte mit ausschließlich nicht-personenbezogenen Daten in der Regel einfacher nutzbar sind. Es empfiehlt sich daher, sofern für die konkrete Nutzung ein Personenbezug nicht erforderlich ist, auf die Verarbeitung personenbezogener Daten zu verzichten. Der Personenbezug ist im Datenschutzrecht allerdings (noch) sehr weit zu verstehen, da er auch personenbeziehbare Daten umfasst. Ob ein Datum personenbeziehbar ist, ist gerade bei pseudonymisierten Daten mitunter schwer zu beurteilen. Eine genaue rechtliche Grenzziehung fehlt aktuell noch. Die europäischen Gerichte haben jedoch gewisse Leitlinien festgelegt, die auf die zu beurteilende rechtliche Möglichkeit zur Re-Identifizierung aus Sicht des Datenempfängers abstellen.[34] Diese Rechtsprechung scheint die bislang noch vorherrschende absolute Sichtweise der Datenschutzbehörden aufzuweichen. Ein weiterer Weg, den Personenbezug zu vermeiden, ist die Anonymisierung von Daten, z.B. durch Aggregierung, Generalisierung oder Randomisierung. Allerdings werden auch an die Anonymisierung von Daten hohe Anforderungen gestellt.[35]

Sind personenbezogene Daten für das konkrete Datenprodukt erforderlich oder besteht die rechtliche Möglichkeit zur Re-Identifizierung, sind die datenschutzrechtlichen Regelungen einzuhalten.

Die datenschutzrechtliche Bewertung von Datenprodukten ist geprägt von den Prinzipien des Verbots mit Erlaubnisvorbehalt, der Datensparsamkeit und der Zweckbindung.

Die Verarbeitung der Daten bedarf daher zunächst einer Rechtsgrundlage, die bei Datenprodukten üblicherweise das die Interessen der betroffenen Person überwiegende berechtigte Interesse des Verantwortlichen (Art. 6 Abs. 1 Satz 1 lit. f) DS-GVO) oder die Einwilligung der betroffenen Person (Art. 6 Abs. 1 Satz 1 lit. a) DS-GVO) sein dürfte. Der Vorteil der Rechtsgrundlage des berechtigten Interesses liegt darin, dass die betroffene Person hierauf keinen Einfluss (wie etwa bei dem Widerruf einer Einwilligung) nehmen kann. Der Nachteil ist jedoch, dass für diese Rechtsgrundlage eine Interessenabwägung erforderlich ist, für die nicht immer klare Kriterien vorliegen, was zu einer gewissen Rechtsunsicherheit führt. Die Problematik bei der Einwilligung ist allerdings, dass diese unter Beachtung der entsprechenden formellen Vorschriften einzuholen ist und von der betroffenen Person jederzeit mit Wirkung für die Zukunft widerrufen werden kann. Dies stellt besondere Herausforderungen an das Sammeln von Daten und die Pflege von Datenpools.

Der Grundsatz der Datensparsamkeit besagt, dass personenbezogene Daten dem Zweck angemessen und erheblich sowie auf das für die Zwecke der Verarbeitung notwendige Maß beschränkt sein müssen. Dies steht naturgemäß im Widerspruch zu der Tendenz, eine möglichst umfassende und umfangreiche Datenbasis zu nutzen. Umso wichtiger ist die inhaltliche Datenqualität eines Datenprodukts, um das Prinzip der Datensparsamkeit einhalten zu können. Die verarbeiteten Daten müssen von Bedeutung für das Datenprodukt und seine Anwendung sein.

34 EuGH, Urteil vom 19.10.2016 – Rs. C-582/14 – Breyer; EuG, Urteil vom 26.04.2023 – Rs. T-557/20 - SRB)

35 Positionspapier zur Anonymisierung unter der DSGVO unter besonderer Berücksichtigung der TK-Branche vom 29. Juni 2020, S. 9 f., ↗ https://www.bfdi.bund.de/SharedDocs/Downloads/DE/Konsultationsverfahren/1_Anonymisierung/Positionspapier-Anonymisierung.pdf?__blob=publicationFile&v=4

Sie müssen ferner dem Zweck angemessen sein, d. h. sie dürfen nicht zu tief in die Rechte der betroffenen Person eingreifen. Und schließlich darf kein gleich geeignetes milderes Mittel zur Verfügung stehen, den beabsichtigten Zweck zu erreichen.

Das Datenschutzrecht gebietet es also, genau abzuwägen, welche Daten für ein bestimmtes Datenprodukt wirklich wichtig und notwendig sind.

Die Frage der Zweckbindung ist vor allem im Zusammenhang mit der Rechtsgrundlage sowie den Informationspflichten gegenüber betroffenen Personen von Bedeutung. Werden personenbezogene Daten, die für einen anderen Zweck erhoben wurden, erst später für Zwecke des Datenprodukts genutzt, handelt es sich in der Regel um eine Zweckänderung, für die unter Umständen die bei Erhebung der Daten geltende Rechtsgrundlage nicht mehr einschlägig ist oder die ursprünglichen Informationen nicht ausreichend waren. Es empfiehlt sich daher, bereits bei Erhebung der Daten den Zweck der Nutzung für Datenprodukte bereits vorzusehen (sofern im Übrigen datenschutzrechtlich zulässig).

5.3 Geschäftsgeheimnisse

Datenprodukte können auch Informationen enthalten, die als Geschäftsgeheimnisse zu qualifizieren sind, wenn die weiteren Voraussetzungen des Geschäftsgeheimnisgesetzes erfüllt sind (vgl. § 2 GeschGehG). Ist dies der Fall, so kann dies bei einer rein internen Verwendung möglicherweise vernachlässigt werden. Bei einer externen Verwendung eigener Geschäftsgeheimnisse hingegen kann dies zum Verlust des Schutzes führen, da die Informationen dann unter Umständen frei zugänglich sind und damit nicht mehr die für den Schutz erforderlichen Voraussetzungen nach dem GeschGehG erfüllen. Sollen die geschützten Informationen trotzdem verwendet werden, ist zwingend darauf zu achten, vertragliche, technische und organisatorische Geheimhaltungsmaßnahmen zu ergreifen. Bei einem Vertrieb von externen Datenprodukten kann es unter Umständen auch zur Verletzung von fremden Geschäftsgeheimnissen kommen und Ansprüche des Inhabers des Geschäftsgeheimnisses auf Unterlassung, Vernichtung, Schadensersatz etc. auslösen.

5.4 Weitere geschützte Inhalte

Datenprodukte können zudem weitere geschützte Inhalte beinhalten, z. B. urheberrechtlich geschützte Bilddaten oder geschützte Datenbanken bzw. Datenbankwerke. Auch insoweit ist je nach Art der beabsichtigten Nutzung sicherzustellen, dass die hierfür erforderlichen Rechte bezüglich der geschützten Inhalte vorliegen.

5.5 Datenverträge

Gegenstand von Datenprodukten ist – stark verkürzt – das Sammeln von Daten, deren intelligente Auswertung und das Ableiten nutzbarer Ergebnisse. Der Fokus liegt daher auch rechtlich auf Fragen der Zulässigkeit des Sammelns der Daten, der rechtlichen Befugnis, diese auszuwerten und sie gegebenenfalls für die Umsetzung der Ergebnisse zu verwenden. Hinzu kommt, dass Daten sowohl aus internen als auch aus externen Quellen stammen können, Datenprodukte mithin von Unternehmen selbst erstellt oder von Dritten zur Nutzung (zeitlich beschränkt oder dauerhaft) erworben werden, und für interne und externe Zwecke verwendet werden können.

Um den rechtlichen Regelungsbedarf im Zusammenhang mit Datenprodukten bestimmen zu können, sind – ausgehend vom sog. 3-Ebenen-Modell (Semantik, Syntax, Struktur) – der Inhalt der Daten, ihre technischen Eigenschaften sowie die Art ihrer Verkörperung zu beachten.

Da es – wie erwähnt – ein Dateneigentum oder ein ähnliches Recht (derzeit) nicht gibt, erfolgt die Zuordnung zum einen über die faktisch-technische Kontrolle und zum anderen – meist damit verknüpft – über vertragliche Vereinbarungen, die etwaig entstehende Daten zumindest vertraglich einer Partei zuweisen, die dann üblicherweise auch die faktisch-technische Kontrolle innehat. Dabei ist zu beachten, dass diese vertraglichen Vereinbarungen nur zwischen den Vertragsparteien (»inter partes«), nicht jedoch gegenüber Dritten (»intra omnes«) wirkt.

Die Art der Verkörperung der Daten, mithin die Frage, wer die faktisch-technische Kontrolle über die Daten hat, bestimmt also in der Regel den Ausgangspunkt vertraglicher Vereinbarungen über Datenprodukte. Der Inhaber der faktisch-technischen Kontrolle ist daher immer Partei eines Datennutzungsvertrags. Er ist es, der zunächst vertraglich verpflichtet ist, der anderen Partei den Zugang zu dem Datenprodukt zu ermöglichen.

Vertragsgegenstand

Das Datenprodukt ist der Gegenstand jedes Datennutzungsvertrags. Es ist daher eine Grundvoraussetzung, das Datenprodukt im Vertrag genau zu beschreiben und im Rahmen der Vertragserfüllung zu prüfen, ob das tatsächliche Datenprodukt der vertraglichen Vereinbarung entspricht. Hierzu gehört auch die Art und Weise der Zurverfügungstellung. Die Beschreibung ist abhängig von der Art des Datenprodukts und sollte alle drei Ebenen – Inhalt, Format, Übermittlungsweg – adressieren. Bei feststehenden Datensätzen ist es beispielsweise üblich, nach einer ggf. gemeinsam von den Parteien durchgeführten inhaltlichen Validierung der Daten, das Datenprodukt mittels eines Hashwertes oder einer Prüfsumme eindeutig zu kennzeichnen und dies vertraglich auch zu dokumentieren. So kann der Empfänger des Datenprodukts erkennen, ob das zur Verfügung gestellte Datenprodukt mit dem vertraglich geschuldeten übereinstimmt. Bei dynamischen Datenprodukten kann auf Beispieldatensätze, die Beschreibung der Daten als Output eines spezifischen Erzeugungsprozesses oder auch auf die Beschreibung auf semantischer Ebene zurückgegriffen werden. Auf Syntaxebene kommt es insbesondere darauf an, das Dateiformat zu vereinbaren, in dem das Daten-

produkt zur Verfügung zu stellen ist. Die Beschreibung der Art und Weise der Zurverfügungstellung ist ebenfalls zu regeln (z. B. Remote-Zugriff, Download, Live-Feed). Genauso wichtig ist festzulegen, ob es sich bei den Daten um Rohdaten, strukturierte Daten, semi-strukturierte Daten oder abgeleitete Daten handelt. Je nach Kategorie gelten unterschiedliche Anforderungen an die Datenprodukte und ihre Nutzbarkeit.

Maßgebliche Anforderung an Datenprodukte ist ferner die rechtliche und faktische Datenqualität. Die faktische Qualität betrifft Kriterien wie Aktualität, Granularität, Vollständigkeit und Richtigkeit. Die rechtliche Qualität betrifft etwa Fragen nach Rechten Dritter oder zu datenschutzrechtlichen Anforderungen. Vereinbart werden können hier auch gemeinsame Verfahren und Prozesse zur Überprüfung der Datenqualität.

Werden die rechtlichen Anforderungen nicht eingehalten, drohen Schadensersatzansprüche, Unterlassungsansprüche oder behördliche Maßnahmen, z. B. die Verhängung von Bußgeldern durch die Aufsichtsbehörden.

Bestimmung der berechtigten Nutzer

Hinsichtlich der Nutzungsbefugnis ist zunächst zu regeln, wer ein Datenprodukt nutzen darf. Dies kann beschränkt sein auf den Vertragspartner oder auch dessen verbundene Unternehmen oder sonstige Dritte, z. B. Dienstleister, einbeziehen. Zudem ist eine etwaige Exklusivität zu regeln.

Nutzungsart und Nutzungszweck

Zu regeln sind zudem die Nutzungsarten- und Zwecke. So ist beispielsweise festzulegen, ob das Datenprodukt nur für interne Zwecke des Vertragspartners genutzt werden darf, ob es verändert, kombiniert oder weitergegeben werden darf. Auch spezifische Nutzungszwecke, wie die Nutzung zum Training von KI-Modellen, sollten detailliert geregelt werden. In der Praxis erfolgt insoweit häufig eine Orientierung an den urheberrechtlichen Vertragsgestaltungen. Die Beschreibung der Nutzungsarten- und Zwecke sollte aber unabhängig davon möglichst genau und abschließend sein, da Rechte an Datenprodukten nicht dinglich, sondern nur vertraglich eingeräumt werden. Soweit in dem Datenprodukt geschützte Rechte enthalten sind, sind auch hierzu entsprechende Regelungen vorzusehen. Handelt es sich um Schutzrechte Dritter, sind die Beschränkungen zu beachten, denen derjenige, der die Daten zur Verfügung stellt, selbst unterliegt.

Nutzungsdauer

Zu regeln ist ferner die Nutzungsdauer, d. h. insbesondere, ob es sich um eine dauerhafte Berechtigung zur Nutzung handelt oder eine zeitlich befristete. Dies ist auch ein wichtiges Indiz für die vertragstypologische Einordnung.[36]

36 vgl.. hierzu: Rosenkranzun/Scheufen, Die Lizensierung_ von nicht-personenbezog_enen Daten, ZfDR 2022, S. 159 ff. _.

Gewährleistung und Haftung

Im Rahmen der Gewährleistungsregelungen kommt es darauf an, ob der Vertrag als Kaufvertrag, Mietvertrag oder als Vertrag eigener Art (Vertrag sui generis) einzuordnen ist. Die vertragstypologische Einordnung von datenbezogenen Verträgen ist bislang noch nicht vollständig geklärt. Umso wichtiger ist es, entsprechende vertragliche Regelungen vorzusehen.

Je nach Vertragsgegenstand kann es sich aus Empfängersicht anbieten, bestimmte, ggf. auch mit einer Vertragsstrafe bewährte, Service Level zu vereinbaren. Ferner sollten Regelungen für den Fall getroffen werden, dass ein Datenprodukt nicht den vereinbarten Beschaffenheitskriterien entspricht. Neben Nachbesserungs- oder Nachlieferungspflichten kommen Minderung der Vergütung, Schadensersatz oder Kündigung bzw. Rückabwicklung des Vertrags in Betracht. Zudem sollten Rechtsfolgen für den Fall geregelt werden, dass ein Datenprodukt gegen Schutzrechte Dritter verstößt, Geschäftsgeheimnisse verletzt oder nicht den datenschutzrechtlichen Anforderungen genügt. Die Regelungen sind je nach Interessenlage unterschiedlich auszugestalten. Während derjenige, der die Daten zur Verfügung stellt, ein Interesse daran hat, für ihn nachteilige Rechtsfolgen nach Möglichkeit zu beschränken, hat der Empfänger ein Interesse nach möglichst umfangreichen Zusagen und rechtlichem Schutz bei deren Verletzung. Zudem sind bei Standardbedingungen die Schranken des AGB-Rechts zu beachten. Dies gilt ebenso für Regelungen zur Haftung und etwaige Haftungshöchstgrenzen.

»Schicksal« der Daten

Ebenfalls zu regeln ist das »Schicksal« der Daten. Ist der Gegenstand des Vertrags die möglichst weitreichende Übertragung von Rechten, kann eine Löschung der Daten bei demjenigen, der die Daten zur Verfügung stellt, nach ihrer Übertragung vereinbart werden.

Rechte an abgeleiteten Daten

Da Datenprodukte vielfach dazu genutzt werden, abgeleitete Daten hiervon zu erzeugen oder mittels der Daten neue Ergebnisse zu erzielen, ist es wichtig zu regeln, wer in welchem Umfang berechtigt sein soll, diese Ergebnisse und abgeleiteten Daten zu nutzen.

Vergütung

Ein weiterer wichtiger Punkt ist die Regelung der Vergütung für die Datennutzung und etwaige damit verbundene Dienstleistungen.

Ausblick: Data Act

Mit dem Data Act[37] werden sich die Datenwirtschaft und das Daten(vertrags)recht wandeln. Der Data Act nimmt eine sehr weitreichende Allokation der Rechte an Daten

37 Europäische Kommission. 2022. Vorschlag für eine Verordnung des Europäischen Parlaments und des Rates über harmonisierte Vorschriften für einen fairen Datenzugang und eine faire Datennutzung (Datengesetz). ↗ https://eur-lex.europa.eu/le_al-content/DE/TXT/?uri=CELEX%3A52022PC0068&qid=1678871812636

beim Datenerzeuger (im Data Act: Nutzer datenerzeugender Produkte und Dienstleistungen) vor und gewährt gesetzliche Datenzugangsansprüche, die vertraglich zu regeln sind. Zudem reguliert er vertragliche Gestaltungsmöglichkeiten. Der Data Act ist nicht nur auf nicht-personenbezogene Daten, sondern auch auf personenbezogene Daten anwendbar und stellt so neben der DS-GVO für personenbezogene Daten einen zweiten Rechtsrahmen dar. Die Datenzugangsansprüche nach dem Data Act umfassen auch Geschäftsgeheimnisse, die durch entsprechende Maßnahmen zu sichern sind. Nur in Ausnahmefällen darf der Datenzugang zum Schutz von Geschäftsgeheimnissen (z. B. im Bereich IT-Sicherheit) eingeschränkt oder verweigert werden.

5.6 Fazit

Derzeit sind die Grundlage des Datenrechts zum einen die faktisch-technische Kontrolle als Ausgangspunkt einer Datenzuordnung und vertragliche Vereinbarungen. Da die Diskussion um die vertragstypologische Einordnung von Datenverträgen noch nicht abgeschlossen ist, kommt einer umfassenden Vertragsgestaltung hohe Bedeutung zu.

Die vertraglichen Regelungen sind entsprechend den Interessen der beteiligten Parteien sowie der beabsichtigten Nutzung (intern/extern) zu gestalten. Besonderer Bedeutung kommt dabei der Beschreibung des Vertragsgegenstands, der Datenqualität und der Nutzungsberechtigung zu.

Mit dem Data Act, der für bestimmte Daten eine weitreichende Allokation von Rechten beim Datenerzeuger vornimmt und gesetzliche Datenzugangsansprüche gewährt, werden sich zukünftig noch signifikante Änderungen im Daten(vertrags)recht ergeben.

6 Zusammenfassung

6 Zusammenfassung

In diesem Leitfaden stellen wir – Mitglieder des Bitkom Arbeitskreis Big Data & Advanced Analytics – eine praxisnahe Übersicht zum Entwicklungszyklus für Datenprodukte bereit und diskutieren die Rolle von Governance, Stakeholdern, Datenkatalogen und Datenmarktplätzen dabei. Auch gehen wir auf die Zusammenhänge von diesen Konzepten ein. Wichtige Aspekte der Datenqualität werden dargestellt und in Hinblick auf Automatisierbarkeit diskutiert. Dies komplementieren wir mit einer Skizze der rechtlichen Rahmenbedingungen, welche bei Datenverarbeitung- und Weitergabe relevant sind. Ebenso wichtig wie die theoretische Grundlage sind die Praxisbeispiele, welche diesen Leitfaden weiter illustrieren.

7 Praxisbeispiele

7.1 Einsatzbeispiel Deutsche Telekom

T Systems
DATA INTELLIGENCE HUB

T-Systems International GmbH: Produkte werden in Fabriken hergestellt – warum nicht auch bei Daten endlich nachhaltigen Unternehmenswert generieren?

Steckbrief

↗Telekom Data Intelligence Hub (↗T-Systems International GmbH)

Ausgangslage

In Unternehmen gibt es widersprüchliche Auffassungen zur Datennutzung: Für IT-Abteilungen sind Daten ein Rohstoff – eine Wasserpfütze (»Rohwasser«), während Daten auf Führungsebene eher als Produkt betrachtet werden (Wasserflasche aus dem Regal). So wie niemand seinen Wasserbedarf mit Regenwasser stillen würde, müssen wir Daten veredeln, um skalierbare Mehrwerte zu schaffen.

Herausforderungen

Wenn »Zeit Geld ist«[38], dann ist die Datenanalyse eine wirtschaftliche Katastrophe, denn mehr als 80 Prozent des Zeitbudgets wird mit der Datenverarbeitung- und Veredelung verbracht – nicht mit Ergebnissen. Übliche Produkte werden in Fabriken hergestellt – jetzt brauchen wir Datenfabriken, um den Zeitaufwand zur Generierung von Mehrwert aus Daten zu verringern. Wie würde man eine Datenfabrik aufbauen? Rohdaten gehen hinein, und ein veredeltes Datenprodukt kommt heraus. Aber was passiert dazwischen?

Lösung

In einer Fabrik gibt es üblicherweise eine Reihe verschiedener Abteilungen, welche Abteilungen braucht eine Datenfabrik? Dazu gibt eine Reihe an Erkenntnissen, die sich aus Forschung und Industrie-Best-Practices entwickelt hat (Abbildung 8).

Im Wesentlichen müssen zuerst die Rechte an Rohdaten überprüft werden, bevor Daten gesammelt werden können (Rechte, Lizenzen, Zustimmung der Nutzenden). Dann werden die Daten harmonisiert und richtig beschriftet, damit sie auffindbar sind. Daten müssen auch hinsichtlich ihrer Qualität bewertet werden, denn sonst führt deren Analyse nur zu »Garbage-in-Garbage-out« (GIGO, Qualitätsbewertung). Schließlich sind Governance-Mechanismen erforderlich, um sicherzustellen, dass Daten unter Wahrung der Datenhoheit ausgetauscht werden.

38 Benjamin Franklin. 1748. Ratschläge für junge Kaufleute

Mehrwert

Das Konzept von Datenprodukten ist für viele Unternehmen noch nicht greifbar, daher geht der Telekom Data Intelligence Hub[39] mit realen Beispielen voran: so wie für die Stadt Hamburg mit einem Datenprodukt für intermodales Reisen. Dabei wurde ein künstlicher Travel Agent (Agent) entwickelt und implementiert, bestehend aus drei Software-Elementen (Engines): Ein intermodaler Reiseplanungsrechner (Calculator), ein Personalisierungs-Engine (Matching) und ein Digitaler Zwilling mit Endnutzerprofilen (Profiler). Dadurch wurde deutlich, dass durch Dataspaces neue Anwendungen entstehen, die den Endkundinnen- und Kunden ein bedarfsgerechteres, flexibleres und schnelleres Produktangebot liefern als heute.

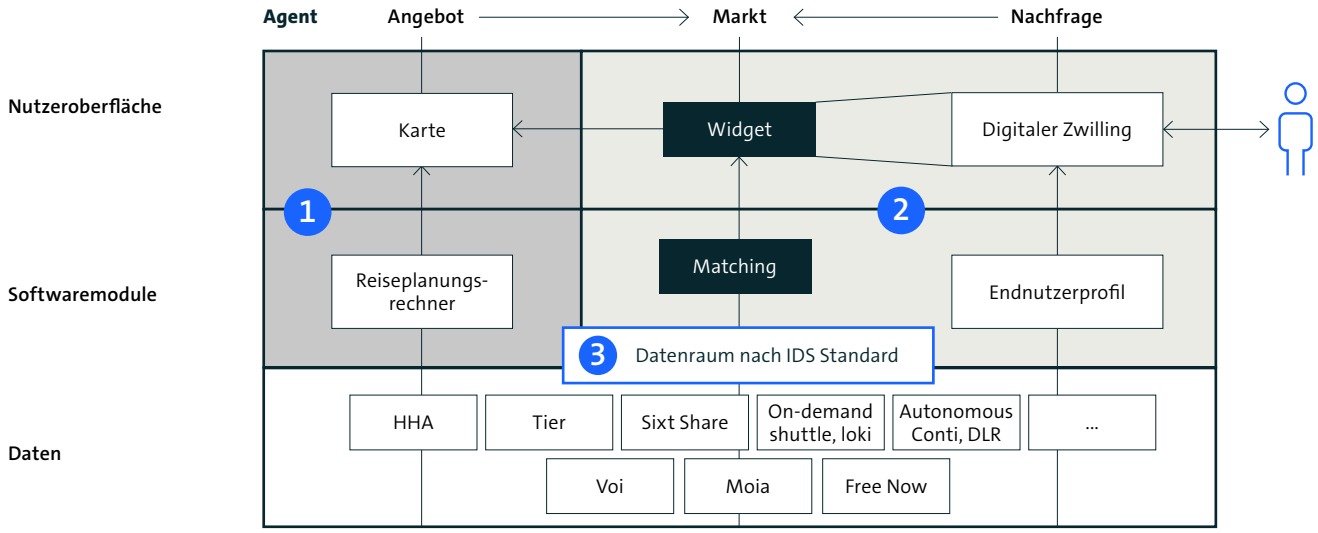


Abbildung 8 – App-Architektur: Intermodales Reisen in Hamburg[40]

Empfehlungen

- Neuartige Technologien wie Dataspaces und Datenfabriken haben keinen Selbstzweck, sie müssen Mehrwerte schaffen. Ein Datenprodukt ist also nicht ausreichend für eine nachhaltige Wertschöpfung aus Daten. Datenprodukte sind die Basis für Endanwendungen wie die Reise-App in Hamburg.
- Unternehmen müssen beginnen, ihre »Hausaufgaben« bzgl. Datenmanagement zu machen, um schlussendlich von innovativen Ansätzen profitieren zu können.
- Bereiten Sie sich vor für den Dreischritt der Innovation: Verbinden Sie sich mit Daten-Service-Ökosystemen, erschaffen Sie Datenprodukte und bauen Ihre eigenen Business-Apps für Ihren Mehrwert!

Kontakt

Prof. Dr. Christoph S. Langdon

Lead, Data Analytics Executive and Scientist

Telekom Data Intelligence Hub (T-Systems International GmbH)

Christoph.schlueter-langdon@t-systems.com

39 Luettmann, A., and L. Loeffler. 2021. Universities get a digitalization boost: Telekom Data Intelligence Hub as a virtual lab. Telekom Data Intelligence Hub Journal (2021-05-18)

40 Lena Agnesmeyer et al. 2022. Wir verändern Mobilität: Erkenntnisse des Reallabors Hamburg für eine digitale Mobilität von morgen. Reallabor Hamburg (RealLabHH) nach Schlueter Langdon, C. 2021b. Agent System with »Bring Your Own Membership« Matchmaking. Working Paper (WP_DCL-DruckerCGU_2021-03), Drucker School of Management, Claremont Graduate University, CA

7.2 Einsatzbeispiel Swisslog / GP+S Consulting

Konzeption datenbasierter Services für die Lagerlogistik: Identifikation datenbasierter Services und Definition einer Umsetzungsroadmap

Steckbrief

Swisslog AG
Swisslog ist führend auf dem Gebiet der daten- und robotergesteuerten automatisierten Logistiklösungen und bietet zuverlässige, modulare Servicekonzepte an. Swisslog beschäftigt mehr als 2.400 Mitarbeitende an über 20 Standorten weltweit und ist Teil der KUKA Gruppe mit einem Jahresumsatz von über 3.2 Milliarden Euro.

Ausgangslage

Swisslog ist als Integrator von Automatisierungslösungen in der Lagerlogistik für die Steuerung von komplexen Logistikprozessen seiner Kunden verantwortlich. Um die Kunden bestmöglich zu unterstützen und den Service kontinuierlich zu verbessern, sollen die verarbeiteten Prozessdaten besser genutzt und das digitale Leistungsportfolio ausgebaut werden. Bei der Identifikation und Bewertung datenbasierter Geschäftsmodelle soll unmittelbares Kundenfeedback und der spätere Kundennutzen im Mittelpunkt stehen. Als global tätiges Unternehmen ist es für die Swisslog wichtig, regionale Unterschiede und Kundenanforderungen zu berücksichtigen. Im Rahmen des Projektes sollen daher alle regionalen Markt- und Produktexperten in einem internationalen und interdisziplinären Projektteam eingebunden werden.

Herausforderungen

Eine zentrale Herausforderung bestand in der Identifikation von datenbasierten Services, die ein akutes Bedürfnis beim Kunden adressieren und auf eine starke Nachfrage treffen. Dabei sollten die Services einerseits über ein gewisses Skalierungspotenzial verfügen, andererseits mussten auch die individuellen Gegebenheiten der Kunden in Bezug auf deren Logistiksystem und die verfügbare Datenbasis berücksichtigt werden.

Eine zusätzliche Herausforderung lag in der Notwendigkeit, die für die datenbasierten Services erforderlichen Fähigkeiten zu entwickeln. Basierend auf dem bestehenden Fähigkeits-Profil der Swisslog, musste die Identifizierung, sowie der quantitative und qualitative Aufbau der benötigten Fähigkeiten erfolgen. Dies war notwendig, um eine Realisierung der geplanten Datenprodukte sicherstellen zu können.

Lösung

Zielbild

- Konsolidierung der bestehenden Strategie, der Marktwahrnehmungen und der bereits bestehenden datenbasierten Services
- Entwicklung eines zukünftigen kundenzentrierten Portfolios für datenbasierte Dienstleistungen

- Zusammenfassung von Anforderungen und Empfehlungen für die erfolgreiche Umsetzung im Unternehmen
- Konsolidierung und Präsentation der Ergebnisse und Empfehlungen für die Entscheidungsfindung über das weitere Vorgehen bei datenbasierten Dienstleistungen

Lösung

- Entwicklung eines am **Design Thinking-Prozess angelehnten Projektvorgehens**, bei dem die Bedürfnisse der Kunden in den Mittelpunkt gestellt werden
- Einbindung eines **interdisziplinären Projektteams**, das alle notwendigen Perspektiven (insb. Markt, Kunde, Produkt, Service, Software) einbringt und das Engagement des gesamten Unternehmens fördert
- Durchführung von **Kundeninterviews**, um deren Herausforderungen zu verstehen und die Erwartungen an datenbasierte Services aufzunehmen.
- **Identifikation** und Beschreibung **von datenbasierten Service-Ideen**, inkl. Kalkulation einer Potenzialabschätzung
- Identifikation der erforderlichen Fähigkeiten für eine Realisierung datenbasierter Services mit Hilfe der Methodik des **Business Capability Mapping**
- Reifegradbewertung der existierenden Fähigkeiten, Aufzeigen konkreter Handlungsbedarfe und Überführung in eine **Umsetzungsroadmap**

Mehrwert

- Ein vom gesamten Management getragenes **Ziel-Portfolio datenbasierter Services** im Kontext Lagerlogistik, inklusive Beschreibung und Potenzialabschätzung
- **Direktes Kundenfeedback** zu den erarbeiteten Geschäftsmodellansätzen
- **Transparenz über die erforderlichen** technischen und unternehmerischen **Fähigkeiten**, sowie deren heutiger Reifegrad im Unternehmen, visualisiert in einer Capability Map
- Eine mit allen relevanten Stakeholdern abgestimmte **Umsetzungsroadmap** zur Realisierung datenbasierter Services

Empfehlungen

- Vom Kunden und dessen Bedürfnissen, sowie dessen Ausgangssituation (Logistiksystem und verfügbare Datenbasis) denken und diesen frühzeitig in das Vorhaben einbinden
- Gemeinsamkeiten zwischen den Anforderungen ermitteln, um eine Skalierbarkeit und effiziente Umsetzung der datenbasierten Services zu realisieren
- Konkrete datenbasierte Services identifizieren und diese mit potenziellen Kunden im direkten Dialog verproben
- Ein strukturiertes Vorgehen aufsetzten, mit dem greifbare Ergebnisse in einem überschaubaren Zeithorizont erzielt werden
- Die Unterstützung des gesamten Top-Managements sicherstellen, um die Initiative auch gegen interne Widerstände voranbringen zu können
- Eine niederschwellige Einbindung der relevanten Entscheider und Fachexperten ermöglichen
- Beim internen Aufbau neuer, datenbasierter Fähigkeiten einen Fokus setzen, auf solche mit strategischer Bedeutung und dem Potenzial einen Wettbewerbsvorteil zu erzeugen
- Konkrete Umsetzungsmaßnahmen ableiten und Verantwortlichkeiten festlegen

Kontakt

Felix Ruscheweyh

Head of Corporate Program Management

Swisslog AG

felix.ruscheweyh@swisslog.com

Marc Schumacher

Senior Consultant

GP+S Consulting GmbH

marc.schumacher@gps-consulting.com

7.3 Einsatzbeispiel PwC Deutschland

Die Business Driver Database – eine zentrale Anlaufstelle für standardisierte und geprüfte makroökonomische Daten bei PwC

Steckbrief

PricewaterhouseCoopers GmbH Wirtschaftsprüfungsgesellschaft

- Wirtschaftsprüfungsgesellschaft, Chief Data Office
- Das Chief Data Office koordiniert die Arbeit mit Daten für effizientere Prozesse und um PwC und ihre Kunden zu fördern

Ausgangslage

Das Chief Data Office von PwC befähigt und unterstützt die Fachbereiche des Unternehmens bei ihren Kundenprojekten mit Expertise und datengetriebenen Lösungen. Vor der Einführung der Business Driver Database (BDD) hatte ein interner Kunde Verbesserungsbedarf bei Zeitaufwand für Recherche, Beschaffung und Aktualisierung von benötigten Daten. Diese Prozesse kosteten Zeit, die für die eigentlichen Kundenprojekte im Bereich Forecasting as a Service-Lösung benötigt wurde. Um dies zu lösen, brauchte es ein zuverlässiges Datenprodukt für einen einfachen und schnellen Zugriff auf umfassende, standardisierte und aktuelle makroökonomische Datensätze.

Herausforderungen

Die mangelnde Zentralisierung und Vereinheitlichung der Integration von Datenanbieterdaten, die notwendige Prüfung der Datenqualität sowie die Dokumentation der bestehenden Daten führten zu einem unübersichtlichen und komplizierten Umfeld und kosteten die Nutzerinnen und -nutzer viel Zeit.

Zudem wurde eine grundsätzliche Expertise benötigt, um die Prüfungen durchzuführen, die beim Chief Data Office angeboten werden. Darüber hinaus führt eine Fragmentierung der Beschaffungs- und Prüfprozesse von makroökonomischen Daten zusätzlich zu Effizienz- und Qualitätseinbußen.

Lösung

Die Vision war es, eine umfassende Quelle für makroökonomische Daten anzubieten. BDD bietet diese Lösung und nimmt die Rolle als zentrale Anlaufstelle ein. Dabei senkt sie Kosten, indem redundante Aufgaben abgebaut werden, beispielsweise durch die zentrale Integration von Datenanbietern. Vorher wurden teilweise die gleichen Daten von externen Anbietern für verschiedene digitale Lösungen bezogen. Durch eine zentrale Stelle werden doppelte Beschaffungen vermieden und die Daten vorab standardisiert und qualitätsgeprüft bereitgestellt. Datensätze von verschiedenen Anbietern werden so systematisch integriert und in einer gemeinsamen Struktur zusammengeführt, um sie den internen Datennutzerinnen und -nutzer zugänglich zu machen.

Die Daten können als solche oder in Verbindung mit anderen verfügbaren Kundendaten maschinell verwertbar für belastbare Analysen genutzt werden. So kann das volle Potenzial makroökonomischer Daten ausgeschöpft und den internen Teams dabei geholfen werden, sich ganz auf ihre Kundenprojekte zu fokussieren. Bei den verschiedenen Anforderungen der Nutzerinnen und -nutzer bei PwC war es wichtig, die Balance zu schaffen zwischen einem vielseitigen Datenprodukt und maßgeschneiderten Lösungen. Dabei stand die Skalierbarkeit und langfristige Vision des Produkts im Mittelpunkt, um möglichst viele Anforderungen abzudecken.

Mehrwert

Die BDD ist eine umfassende Datenbank für makroökonomische Datensätze aus internationalen Datenquellen. Dadurch ergeben sich für die Nutzerinnen und -nutzer dieser Drittanbieterdaten folgende wertvolle Vorteile:

- Bequeme Handhabung standardisierter und harmonisierter makroökonomischer Daten
- Effizienzsteigerung innerhalb der Teams in Bezug auf Zeit- und Kostenoptimierung
- Präzise Prognosen und frühzeitige Warnungen aufgrund der Verfügbarkeit aktueller und umfassender Datensätze
- Gewährleistung eines hohen Maßes an Datenqualität für Analysen
- Einfache Visualisierung von makroökonomischen Daten mit Tools wie PowerBI und SQL

Empfehlungen

Die Übertragbarkeit von Lösungen beachten:
Ein Datenprodukt sollte den Anforderungen der Nutzerinnen und -nutzer entsprechen. In manchen Fällen können einzelne Teile des Datenprodukts für andere Anwendungen genutzt werden. Allerdings stellen verschiedene Nutzergruppen unterschiedliche Anforderungen an die Datenverarbeitung, die Visualisierung und die Datenqualität.

Datenanbieter mit Bedacht wählen:

- Die Geschäftsbedingungen des Datenanbieters können das Teilen oder Übertragen von Daten aus dem Datenprodukt einschränken.
- Je mehr Datenanbieter in ein Produkt integriert werden sollen, desto mehr Anpassungen und Standardisierungen müssen vorgenommen werden (Dokumentation, Ausfälle, Änderungen).

Verantwortlichkeit für das Datenprodukt:
Wie bei jedem anderen Produkt muss sich der Product Owner auch bei Datenprodukten um alle Aspekte der Produktverwaltung kümmern.

Erfolgsfaktoren:

- Ein holistischer Ansatz bei der Entwicklung des Datenprodukts ermöglicht es, die Anwendung jeweils auf individuelle Nutzerbedürfnisse anzupassen und zu übertragen
- Die Compliance-Prüfungen der einzelnen Datenanbieter sind essenziell. Jeder Datenanbieter hat spezifische Nutzungsbedingungen, daher ist besondere Sorgfalt geboten, wenn Daten für kommerzielle und nichtkommerzielle Aktivitäten geteilt werden sollen.

Kontakt

Adel Anwar
Associate, Chief Data Office,
PwC Deutschland
adel.anwar@pwc.com

Marcus Hartmann
Partner und Chief Data Officer,
PwC Deutschland und Europa
marcus.hartmann@pwc.com

Umair Usman
Senior Associate, Chief Data
Office, PwC Deutschland
umair.usman@pwc.com

7.4 Einsatzbeispiel Deutsche Bahn

SAP-Datenprodukte für die Data Management Platform

Steckbrief

DB Systel GmbH, Einheit Maintenance Asset Solutions (MAS)
Systematische Digitalisierung ist der entscheidende Schlüssel zur Starken Schiene. Gemeinsam mit allen Geschäftsfeldern der Bahn und verbundübergreifend arbeiten wir an der Digitalisierung der Geschäftsprozesse und digitalen, ganzheitlichen Lösungen, um exzellente Kundenerfahrungen mit starkem Kundennutzen zu schaffen.

Ausgangslage

Viele Digitalisierungsprojekte der Deutschen Bahn benötigen Daten aus SAP-Systemen. Zum Beispiel sind Verschleißanalysen in der Instandhaltung nicht möglich, ohne Sensordaten vom Zug mit Informationen zu Zugkomponenten und Instandhaltungshistorie aus operativen oder analytischen SAP-Systemen anzureichern. Die Datenversorgung aus SAP erfolgt heute meist durch projektspezifische Punkt-zu-Punkt-Verbindungen – mit schlechter Time-to-Market, hoher Anzahl an redundanten Schnittstellen und wenig Wiederverwendung. Die Umsetzung derartiger Schnittstellen ist oft Release-gebunden und bedarf somit einer langfristigen Planung sowie Konzeption, was im agilen Projektkontext jedoch stellenweise schwierig ist.

Herausforderungen

- Schlechte Time-to-Market bei Datenlieferungen aus SAP-Systemen
- Hohe Anzahl redundanter Schnittstellen, bei niedrigem Wiederverwendungsgrad
- Schnittstellen-Erstellung- und Anpassung ist in der Regel Release-gebunden
- Tiefgreifendes Verständnis von SAP-Datenstrukturen notwendig, um die richtigen Daten aus dem richtigen SAP-System für den jeweiligen Use-Case zu identifizieren
- Keine Standardisierung der Datenversorgung von Digitalisierungsprojekten mit SAP-Daten
- Aufwändige Prozesse hinsichtlich Data Governance, manuelles Erstellen von Schnittstellenbeschreibungen- und Vereinbarungen ohne transparente Lineage

Lösung

SAP-Daten sollen für alle Digitalisierungsprojekte schnell und einfach auffindbar und nutzbar gemacht werden.

Den Kern der Lösung bildet die Kombination des Data Brokers auf Basis von AWS S3 mit dem SAP-Standardtool »Data Intelligence«(DI). SAP DI stellt die Konnektivität zu den verschiedenen SAP-Systemen im Konzern her. Dabei erfolgt vorrangig eine Konnektivität gegen das analytische BW-System und nur im Ausnahmefall gegen das operative ERP-System. Dadurch wird die Schnittstellen-Last der operativen SAP-Systeme sowie die redundante Ausleitung von Daten aus diesen Systemen reduziert.

Beim Einsatz von über 20 unterschiedlichen SAP-Systemen innerhalb des DB-Konzerns werden durch den zentralen Einsatz von SAP DI langfristig Wartungs- und Entwicklungskosten eingespart.

Die Aufbereitung der SAP-Daten zu generischen Datenprodukten erfolgt durch sogenannte Pipelines in der Modeler-Komponente von SAP DI, die die Daten entweder als Vollabzug oder im Delta-Verfahren in der File Drop Zone des Data Brokers bereitstellt. Das Design der Datenprodukte erfolgt dabei in enger Kooperation zwischen IT und Fachanwendern/Data Ownern. Je SAP-Quellsystem wird eine Arbeitsgruppe etabliert, die Datenprodukte fachlich generisch definiert und für die technische Qualitätssicherung und Abnahme sorgt. Die Berechtigungssteuerung erfolgt am Data Broker metadatenbasiert kundenindividuell.

Der Data Broker indiziert die bereitgestellten Daten und stellt diese im Zusammenspiel mit dem DB Data Catalog und dem Business Hub Connect an Abonnenten zur Verfügung. Dabei sichert der DB Data Catalog nicht nur die Auffindbarkeit der publizierten Datenprodukte in einem zentralen Metadata-Repository, sondern bildet auch die Basis einer durchgängigen Data Governance. Hier können Fachanwender toolgestützt Data Contracts mit den jeweiligen Data Ownern schließen, sodass in der Zielausprägung nur wenige Klicks zwischen dem Erkennen eines Datenbedarfs und der Bereitstellung des Datenprodukts notwendig sind.

Mehrwert

- Teilautomatisierte Bereitstellung von SAP-Daten inkl. Freigabeprozessen
- Schnelle Auffindbarkeit bestehender Datenprodukte über den DB Data Catalog auch ohne tiefgreifendes SAP-Verständnis
- Verkürzte Time-to-Market
- Wiederverwendbarkeit der technologischen Lösung für diverse SAP-Verfahren im Konzern, dadurch insgesamt effizientere Wartung und Weiterentwicklung
- Durchgängige Governance und Transparenz der Datenflüsse über verschiedene SAP- und non-SAP-Systeme hinweg
- Ermöglichen von Lineage-Analysen durch konsequente Bereitstellung von Metadaten

Empfehlungen

Die Transformation von Punkt-zu-Punkt Schnittstellen hin zu generischen Datenprodukten ist nicht nur ein technologischer Wandel, sondern bedingt auch einen kulturellen und organisatorischen Change. Daher ist es wichtig, frühzeitig in einem crossfunktionalen Team aus Data Ownern, Entwicklern, Architekten und Anwendern bestehende Prozesse kritisch zu hinterfragen und gemeinsam Automatisierungspotenziale aufzuzeigen. So können frühzeitig bürokratische Hürden abgebaut und Vertrauen in innovative Lösungen gebildet werden. Technologisch existieren diverse Lösungen zur Ausleitung von SAP-Daten, insb. in Multi-Cloud Landschaften. Wir haben Nähe zum SAP-Standard und Capabilities zur metadatenbasierten Datenextraktion in den Mittelpunkt gestellt, um auf zukünftige SAP Releasewechsel und Migration auf S4/HANA vorbereitet zu sein und perspektivisch Lineage-Analysen vom Dashboard bis zurück in SAP-Tabellen zu ermöglichen.

Kontakt

Dr. Marcel-Philippe Breuer

SAP BI Berater

DB Systel GmbH

Philippe.Breuer
@deutschebahn.com

Dr. Daniel Pöppelmann

SAP-Architekt

DB Systel GmbH

daniel.poeppelmann
@deutschebahn.com

7.5 Einsatzbeispiel CRIF

Entwicklung eines Scoreprodukts zur Bonitätsschätzung von Unternehmen unter Berücksichtigung der Datenverfügbarkeit

Steckbrief

CRIF unterstützt Unternehmen & Finanzinstitute ganzheitlich beim Management ihrer Digital Customer Journey mit integrierten B2B2C Identity-, Credit Risk- und Fraud Prevention-Lösungen aus einer Hand.

CRIF Deutschland gehört zur weltweit tätigen CRIF-Gruppe mit Hauptsitz in Bologna. 40+ Länder, 85+ Unternehmen, 6.500+ Experteninnen und Experten.

Ausgangslage

Neben klassischen Datenprodukte, die das Kerngeschäft von CRIF bilden, werden auch insbesondere Scoreprodukte angeboten. In der Vergangenheit waren diese jedoch aufwändig in der Aktualisierung, die Anzahl der verknüpften Datenquellen überschaubar und gerade die Informationsdichte über kleinere Firmen verbesserungswürdig. Industriestandard war, dass Scores trotz sich ändernder wirtschaftlicher Bedingungen 7 und mehr Jahre betrieben wurden.

Herausforderungen

In einer sich immer schneller verändernden Welt ist es wichtig, die Vorhersagequalität von Daten- und Scoreprodukten gerade auch im Zusammenhang mit unerwarteten geopolitischen Ereignissen zu verbessern, und durch regelmäßige Justierung zu erhalten. Außerdem müssen moderne Modelle bei externen Schocks zeitnah anpassbar sein. Dies darf nur minimale Auswirkungen auf die Abläufe der Kunden haben, die bei einem immer höheren Grad an Digitalisierung und Automatisierung, möglichst vollumfängliche Informationsabdeckung in Echtzeit erwarten.

Dabei sind selbstverständlich alle gesetzlichen Rahmenbedingungen der DSGVO ebenso zu erfüllen, wie die regulatorischen Anforderungen in den Branchen, in denen unsere jeweiligen Kunden tätig sind. Als universeller Informationsdienstleister stellt dies höchste Herausforderungen an unser technisches, fachliches und rechtliches Know-How.

Lösung

Ziel war es ein Daten-/Scoreprodukt zu entwickeln, welches, segmentiert nach Datenverfügbarkeit, eine akkurate Einschätzung der Bonität einer wirtschaftlichen Unternehmung unabhängig von der Größe der jeweiligen Unternehmung vollautomatisch bereitstellt.

Dazu wurden Daten aus verschiedenen Quellen zusammengeführt und neben klassischen öffentlichen Registern weitere Datenquellen erschlossen. Wirtschaftliche Unternehmungen wurden nach Datenverfügbarkeit segmentiert und für alle 6 Segmente wurden die 10 – 13 relevantesten Inputs aus mindestens 4 Quellen identifiziert. So wurde sichergestellt, dass den Kunden jederzeit eine State-of-the-Art Vorhersage der Bonität ihrer Geschäftsbeziehungen zur Verfügung steht.

Durch eine entsprechende Logik ist sichergestellt, dass unseren Kunden stets eine optimale Entscheidungsgrundlage zur Verfügung gestellt wird, welche für sie die Grundlage datengetriebener Entscheidungen darstellt. Durch andauerndes Monitoring wird die Vorhersagequalität fortlaufend überwacht und bei Bedarf angepasst, wobei die Skalierung stets so erfolgt, dass bei Kunden keine Anpassungen notwendig sind.

Mehrwert

Eine optimierte Vorhersagequalität mit einer Trennschärfe von 61 Ginipunkten führt zu reduzierten Zahlungsausfällen. Die automatisierte Bereitstellung der Scoreprodukte hilft, Prozesskosten im Griff zu halten.

Sollte im Zuge des Monitorings eine Neukalibrierung des Scores erfolgen, so ist durch die Anpassung der Skalierung sichergestellt, dass dem Kunden keine weiteren Kosten durch Systemanpassungen entstehen. Gleichzeitig bleibt die Trennschärfe konstant.

Das Scoreprodukt ermöglicht es Kunden mit verschiedensten Erfahrungshorizonten eine verlässliche und datengetriebene Bewertung als Entscheidungsgrundlage zu nutzen. Ein objektives Bewertungsverfahren verringert zudem die Gefahr einer subjektiven Entscheidung beim Kunden.

Empfehlungen

- Eine hohe Datenqualität ist essenziell, um dem Kunden eine zuverlässige Entscheidungsgrundlage bieten zu können
- Eine passende technische Infrastruktur ermöglicht eine agile Anpassung an wirtschaftliche Verhältnisse
- Die richtige Auswahl der Daten ist entscheidend, um nicht nur eine maximale Trennschärfe zu erreichen, sondern auch den Anforderungen an den Datenschutz und weiteren Regularien zu gewährleisten.
- Durch eine sorgfältige und objektive Bewertung des Ausfallrisikos kann das Vertrauen von Geschäftskunden gestärkt werden.

Kontakt

Dr. Kerstin Schmidt
Data Scientist
CRIF GmbH
k.schmidt@crif.com

Thorn Thaler
Senior Data Scientist
CRIF GmbH
t.thaler@crif.com

Andreas Kulpa
Executive Director Product Management
CRIF GmbH
a.kulpa@crif.com

8 Mitwirkende

Marcel Altendeitering

Wissenschaftlicher Mitarbeiter Datenwirtschaft
Fraunhofer-Institut für Software- und Systemtechnik ISST

Marcel Altendeitering arbeitet seit 2018 am Fraunhofer ISST und forscht im Rahmen seiner Tätigkeit zu den Themen Datenqualität, Data Spaces und Data Engineering.

Adel Anwar

Associate, Chief Data Office
PwC Deutschland

Adel ist Associate im Chief Data Office und spezialisiert in digitaler Transformation mit Schwerpunkt auf Datenprodukten und Drittdaten.

Stephan Bautz

Senior Manager
PwC Deutschland

Stephan Bautz ist langjähriger Digital Architect und begleitet seine Kunden von der (Daten-)Strategie bis zur Umsetzung.

Dr. Marcel-Philippe Breuer

SAP BI Berater
DB Systel GmbH

Dr. Marcel-Philippe Breuer ist ein erfahrener SAP BI Berater und beschäftigt sich mit der Konzeption sowie Entwicklung komplexer Lösungen im Umfeld von SAP Analytics und den Schnittstellen zu Nicht-SAP-Systemen auf Basis von SAP BW, SAP HANA, SAP Data Intelligence.

Chris Buchhold

Specialist Data Architecture, Data Office
Deutsche Bahn Vertrieb

Chris Buchhold bringt seine umfassende und branchenübergreifende Expertise aus dem Daten Management in das Data Office für den Vertrieb der Deutschen Bahn ein, um die Datenstrategie weiterzuentwickeln und stellt durch die Etablierung einer Data Governance die effektive Bereitstellung und Nutzung von Datenprodukten sicher.

Lukas Feuerstein

Senior Manager | Artificial Intelligence & Data
Deloitte Consulting GmbH

Lukas Feuerstein ist Senior AI & Data Strategy Manager bei Deloitte Consulting mit Fokus auf AI & Data Business Cases, Analytics Operating Model Design und Data-Driven Decisioning.

Marcus Hartmann

Partner und Chief Data Officer
PwC Deutschland

Marcus Hartmann ist Partner bei PwC Deutschland und Chief Data Officer für PwC Deutschland und Europa. Als ausgewiesener Datenexperte hat er seine gesamte Karriere in der Data & Analytics-Industrie verbracht und Unternehmen dabei unterstützt, sich leichter und schneller in einer zunehmend von Daten geprägten Welt zu bewegen.

Dr.-Ing. Ibrahim Halfaoui

AI Expert / Consultant
TÜV SÜD Digital Service GmbH

KI-Experte mit Promotion im Bereich Deep Learning und ein MBA mit Schwerpunkt Innovation. Ich habe mehr als zehn Jahren Erfahrung in verschiedenen Industrien (Automotive, Versicherung, TIC). Bei TÜV SÜD seit November 2022 verantwortlich für das Thema KI-Qualität.

Pascal Hess

Senior Manager & Competence Lead, Data & Analytics, Consulting
Fujitsu Services GmbH

Pascal Hess leitet die Data- und Analytics-Kompetenz und berät Kunden zur nachhaltigen, zielgerichteten und wertstiftenden Nutzung von Daten und Informationen.

Dr. Marvin Jagals

Data Intelligence Center, TOA
Deutsche Bahn AG

Meine Rolle im Konzern: Teil des Workstreams »Data Governance« im House of Data (TOA) bei der Deutschen Bahn AG. Wir als Workstream »Data Governance« treiben die Themen Datenorganisation, Datenqualität und Datenkompetenz aus der Konzernleitung heraus.

Ralph Kemperdick

RaKeTe-Technology

Ralph Kemperdick hat umfangreiche Erfahrung in der Nutzung und Aufbereitung von Daten für die Unternehmenssteuerung mit Schwerpunkt auf Data Engineering, Analytics sowie Künstliche Intelligenz im Cloud-Umfeld.

Dr. Michael Kraus

Rechtsanwalt, Partner, Fachanwalt für Informationstechnologierecht
CMS Hasche Sigle

Dr. Michael Kraus berät und vertritt Unternehmen in allen Fragen des IT- und Datenrechts – von der Gestaltung und Verhandlung von Verträgen zu Technologie- und Digitalisierungsprojekten bis zu den IT-rechtlichen Aspekten von Zukunftsthemen, etwa in den Bereichen KI, Connected Cars, Internet of Things.

Andreas Kulpa

Executive Director Product Management
CRIF GmbH

Datengetriebene Geschäftsmodelle sind seit über 20 Jahren Gegenstand der Tätigkeiten von Andreas Kulpa. Von Big Data über Advanced Analytics bis hin zu Artificial Intelligence

Dr. Till Luhmann

Head of Corporate Development
BTC Business Technology Consulting AG

Dr. Till Luhmanns Arbeitsschwerpunkte liegen in der Strategieentwicklung und der Zukunftssicherung und Innovation, darüber hinaus liegt ein Fokus auf Fragestellungen der zukünftigen nachhaltigen Energieversorgung.

Dr. Michael Pauly

Senior Consultant Marketing Strategy & Analytics
Deutsche Telekom Geschäftskunden GmbH

Dr. Michael Pauly ist als Senior Consultant im Bereich Marketing Strategy & Analytics des Geschäftskundenmarketings der Deutschen Telekom verantwortlich für die Positionierung und Weiterentwicklung von datengetriebenen und performanceorientierten Marketingansätzen.

Nina Popanton

Business Development Executive, Data Intelligence Hub
T-Systems International GmbH

Nina Popanton gestaltet als Business Development Executive bei T-Systems International den Ausbau strategischer Geschäftsfelder und treibt die Skalierung innovativer Dataspace-Lösungen voran. Mit einem klaren Fokus auf nachhaltige digitale Transformation entwickelt sie Lösungsansätze für das Partner:innen- und Kund:innen-Netzwerk von T-Systems.

Dr. Daniel Pöppelmann

SAP BI Architekt
DB Systel GmbH

Dr. Daniel Pöppelmann ist ein Senior Architekt für SAP Analytics und erarbeitet u. a. Lösungen für den Datenaustausch zwischen SAP- und non-SAP Systemen in hybriden Cloud-Architekturen.

Felix Ruscheweyh

Head of Corporate Program Management
Swisslog GmbH

Felix Ruscheweyh verantwortet das Corporate Program Management bei Swisslog und treibt die Entwicklung eines datengetriebenen und kundenzentrierten Serviceportfolios voran.

Prof. Dr. Chris Schlueter Langdon

Business Lead Data Intelligence Hub
Catena-X Product Manager
T-Systems International GmbH

Chris ist Professor für Data Sciences, Business Lead des Telekom Data Intelligence Hubs, der primären Dataspace-Einheit des Telekom Konzerns; und in dieser Funktion ist er agiler Produktmanager von Catena-X, dem offenen Datenökosystem für Automotive, und Konsortialführer von Gaia-X für Advanced Mobility Services.

Dr. Kerstin Schmidt

Data Scientist
CRIF GmbH

Dr. Kerstin Schmidt arbeitet als Data Scientist und beschäftigt sich hauptsächlich mit der Erstellung von Bewertungsmodellen für Unternehmen.

David Schönwerth

Bereichsleiter Data Economy
Bitkom e. V.

David Schönwerth verantwortet die Themenbereiche Datenpolitik, Datenwirtschaft und Data Management beim Digitalverband Bitkom.

Theresa Schramm

Consultant Marketing Strategy & Analytics
Deutsche Telekom Geschäftskunden GmbH

Theresa Schramm ist aktiv im Bereich Marketing Strategy & Analytics des Geschäftskundenmarketings der Deutschen Telekom und verantwortlich für das operative Performance Marketing sowie die Positionierung und Weiterentwicklung von datengetriebenen Marketingansätzen.

Marc Schumacher

Senior Consultant
GP+S Consulting GmbH

Marc Schumacher beschäftigt sich mit der Entwicklung datenbasierter Geschäftsmodelle und begleitet Unternehmen von der Entwicklung eines strategischen Zielbildes bis zur konkreten Umsetzung datenbasierter Services.

Dr. Michael Stadler

Unternehmensentwicklung, Corporate Development, Sen Business Dev Man
BTC Business Technology Consulting AG

Michael Stadler beschäftigt sich als Informatiker und Business Developer seit 15 Jahren für verschiedene Branchen mit Digitalisierungsthemen.

Thorn Thaler

Senior Data Scientist
CRIF GmbH

Seit über 10 Jahren im Bereich Machine Learning, Modellierung und klassischer Data Science tätig, arbeitet Thorn Thaler heute bei der CRIF als Senior Data Scientist.

Dr. Adam Trendowicz

Senior Data Scientist, Abteilung Data Science
Fraunhofer-Institut für Experimentelles Software Engineering IESE

Derzeit liegt der Tätigkeitsschwerpunkt von Dr. Trendowicz auf Datenqualität- und Vorbereitung im Kontext von maschinellem Lernen sowie auf dem Lean Deployment von datengetriebenen Innovationen auf Basis von Lösungen aus den Bereichen maschinelles Lernen und Künstliche Intelligenz.

Umair Usman

Senior Associate, Chief Data Office
PwC Deutschland

Umair Usman ist Senior Associate im Chief Data Office von PwC Deutschland und fokussiert sich auf Datenprodukte sowie Third Party Data Enablement innerhalb von PwC Deutschland.

Dr.-Ing. Sebastian Werner

Principal Consultant Data & AI
Thoughtworks Deutschland GmbH

Komplexität aufs nötigste Reduzieren, ungewöhnliche Lösungen finden und Fokus auf nachhaltige Anwendbarkeit legen – all das treibt Sebastian als Datenstratege, Architekt von Dateninfrastruktur und passionierter »Simulant« bei Thoughtworks voran.

Bitkom vertritt mehr als 2.200 Mitgliedsunternehmen aus der digitalen Wirtschaft. Sie generieren in Deutschland gut 200 Milliarden Euro Umsatz mit digitalen Technologien und Lösungen und beschäftigen mehr als 2 Millionen Menschen. Zu den Mitgliedern zählen mehr als 1.000 Mittelständler, über 500 Startups und nahezu alle Global Player. Sie bieten Software, IT-Services, Telekommunikations- oder Internetdienste an, stellen Geräte und Bauteile her, sind im Bereich der digitalen Medien tätig, kreieren Content, bieten Plattformen an oder sind in anderer Weise Teil der digitalen Wirtschaft. 82 Prozent der im Bitkom engagierten Unternehmen haben ihren Hauptsitz in Deutschland, weitere 8 Prozent kommen aus dem restlichen Europa und 7 Prozent aus den USA. 3 Prozent stammen aus anderen Regionen der Welt. Bitkom fördert und treibt die digitale Transformation der deutschen Wirtschaft und setzt sich für eine breite gesellschaftliche Teilhabe an den digitalen Entwicklungen ein. Ziel ist es, Deutschland zu einem leistungsfähigen und souveränen Digitalstandort zu machen.

Bitkom e.V.

Albrechtstraße 10
10117 Berlin
T 030 27576-0
bitkom@bitkom.org

bitkom.org

bitkom