

Data Mesh – Datenpotenziale finden und nutzen

Ein Praxisleitfaden vom
Digitalverband Bitkom

Herausgeber

Bitkom e. V.
Albrechtstraße 10 | 10117 Berlin

Ansprechpartner

David Schönwerth | T 030 27576-179 | d.schoenwerth@bitkom.org

Verantwortliches Bitkom-Gremium

AK Big Data & Advanced Analytics

Autorinnen und Autoren

Sebastian Baumann | DATAbility
Stephan Bautz, Ronny Garz, Goran Loncar | PwC
Ralf Böhme | MT AG
Lukas Feuerstein, Martin Günther | Deloitte Consulting
Lars Francke | Stackable
Carsten Fritsch | Deutsche Bahn
Rüdiger Karl | SAP
Ralph Kemperdick | Microsoft
Niklas Lümme | Eucon Group
Vaibhavi Rege | Allianz Direct
Dr. Nora Reich | KfW
Dr. Gerald Ristow | Software AG
Prof. Dr. Christoph Schlueter Langdon | Deutsche Telekom
Dominik Schneider | Merck Group
David Schönwerth | Bitkom
Dr. Thomas Vollmer | Philips
Dr. Sebastian Werner | Dataiku

Gestaltung

Katrin Krause

Titelbild

Rakicevic Nenad – pexels.com

Copyright

Bitkom 2022 sowie jeweilige Rechteinhaber

Diese Publikation stellt eine allgemeine unverbindliche Information dar. Die Inhalte spiegeln die Auffassung im Bitkom zum Zeitpunkt der Veröffentlichung wider. Obwohl die Informationen mit größtmöglicher Sorgfalt erstellt wurden, besteht kein Anspruch auf sachliche Richtigkeit, Vollständigkeit und/oder Aktualität, insbesondere kann diese Publikation nicht den besonderen Umständen des Einzelfalles Rechnung tragen. Eine Verwendung liegt daher in der eigenen Verantwortung des Lesers. Jegliche Haftung wird ausgeschlossen. Alle Rechte, auch der auszugsweisen Vervielfältigung, liegen beim Bitkom oder den jeweiligen Rechteinhabern.

1	Einleitung	6
2	Evolution der Architekturpattern	7
3	Architektur	10
	Domain-oriented ownership	11
	Data as a Product	11
	Self Service Data Platform und Data Factories	12
	Federated Computational Governance	13
	Zusammenfassung	14
4	Die Bedeutung für Unternehmen	16
5	Organisation, Führung und Kultur im Data Mesh Kontext	18
	Data Governance	18
	Data Stewardship	19
	Data Culture	20
6	Einsatzbeispiele	22
	Einsatzbeispiel Allianz Direct	23
	Einsatzbeispiel DATAbility	25
	Einsatzbeispiel Deutsche Bahn	27
	Einsatzbeispiel Eucon Group	29
	Einsatzbeispiel PwC	31
7	Einführung und Umsetzung	34

1	Die Evolution der Datenarchitekturen	8
2	Logische Architektur	14

- [1] ↗<https://martinfowler.com/articles/data-monolith-to-mesh.html>
- [2] ↗<https://www.bigdata-insider.de/vom-data-lake-zum-data-mesh-a-960160/>
- [3] ↗https://www.youtube.com/watch?v=_bmYXWCxF_Q (Zhamgk Deghani – Intro to DataMesh)
- [4] ↗[Data Mesh defined | James Serra's Blog](#)
- [5] ↗[What is an Azure landing zone? – Cloud Adoption Framework | Microsoft Docs](#)
- [6] ↗[Azure Data Landing Zone template](#)
- [7] ↗[Azure Data & Analytics Landing Zone template](#)
- [8] ↗https://en.wikipedia.org/wiki/Data_mesh
- [9] ↗<https://towardsdatascience.com/what-is-a-data-mesh-and-how-not-to-mesh-it-up-210710bb41e0>
- [10] ↗<https://towardsdatascience.com/data-mesh-101-everything-you-need-to-know-to-get-started-72087f5a7d91>

1 Einleitung

Seien wir ehrlich, wenn wir von einer datengetriebenen Organisation sprechen, meinen wir damit eigentlich eine besser informierte Organisation. Und wenn wir von Organisation sprechen, meinen wir damit eigentlich Menschen und Teams. Wir wollen Menschen näher an die Daten ihrer Organisation bringen. So können sie individuell und im Team besser informierte Entscheidungen treffen. Aber nicht nur das.

Teams, die in Echtzeit auf vielfältige Daten ihrer Organisationen Zugriff haben, können operative und strategische Ideen auf solider Datengrundlage pilotieren. Auf einmal sind schnellere Innovationszyklen möglich, denn langfristige Aufbereitungsprozesse verkürzen sich. Die Auswirkung von vorher nicht (richtig) verfügbaren Informationen auf Produktivität und Profitabilität unterschiedlicher Produkte und Services kann einfacher ausprobiert und genutzt werden.

Obwohl sich wohl nur die wenigsten gegen datengetriebene Prozesse und Geschäftsmodelle aussprechen würden, machen fehlende, veraltete oder qualitativ nicht ideale Daten und Analysemöglichkeiten vielen Initiativen einen Strich durch die Rechnung. Zu aufwändig, zu viel Abstimmung, zu viel Aufbereitung, ungewisse Ergebnisse, zu teuer. Während die Vision einer umfassend datengetriebenen Organisation, von Beginn an oder auch als Transformationsprozess, in der Vergangenheit eher utopisch klang, könnte eine Data Mesh Architektur vielen zu einem Durchbruch verhelfen.

In diesem Leitfaden wollen wir deshalb aus der Praxis für die Praxis einen Überblick zum Thema Data Mesh geben und drei Fragen beantworten:

Was hole ich mir mit Data Mesh ins Haus?

Was kann ich davon erwarten?

Wie kann ein Data Mesh Projekt ablaufen?

Wir streben mit diesem Leitfaden ausdrücklich keine umfassende theoretische Analyse des Themas an, sondern wollen einen möglichst praxisrelevanten Input für Anwender und Interessierte liefern.

2 Evolution der Architekturpattern

Dass die Nutzung von Daten für neue interne und externe Produkte, Reporting und Analytics von vielen Unternehmen vorangetrieben wird, lässt sich Branchen- und größenübergreifend beobachten. Dabei geht es sowohl um historische als auch in die Zukunft gerichtete Auswertungen, für die oft interne und externe Daten aus verschiedenen Unternehmensbereichen kombiniert werden müssen. Über die letzten Jahrzehnte haben sich hierfür verschiedene Arten der Datenmodellierung als Quasistandard der jeweiligen Zeit etabliert. Zu den verbreitetsten gehören dabei das Data Warehouse (»DWH«) und der Data Lake (»DL«). Andere Designmuster wie z.B. Data Factory oder Virtualisierung wurden für bestimmte Spezialfälle eingesetzt, sind aber bei weitem nicht so verbreitet wie DWH und DL.

Lange wurden Daten für Reportingzwecke nach den von Inmon bzw. Kimball definierten Data Warehouse Paradigmen¹ verarbeitet und zur Verwendung vorbereitet. Dafür wurden sie vom operativen System in eine dispositive Datenbank kopiert, welche mit einem Schema definiert war, welches die beabsichtigten Reports bestmöglich unterstützte. Nach dem Beladen der dispositiven Datenbank wurden entsprechende Transformations-, Aggregations- und Datenqualitätsschritte durchgeführt (*ETL – Extract, Transform, Load*), um die Daten entsprechend aufzubereiten und in *Data Marts* zur Verfügung zu stellen. Die Verantwortung für die Aufbereitung der Daten lag dabei beim zentralen Data Warehouse Team.

Mit Aufkommen neuer Analytics Use Cases und neuer technischer Möglichkeiten, etwa durch die Cloud, stiegen auch die Ansprüche der Datennutzer an die Daten, sodass Rohdaten (teilweise in Echtzeit) anstelle von hoch aggregierten und verarbeiteten Daten bereitgestellt werden sollten. Auch sollte ein Teil der Verantwortung zur Durchführung der Transformations- und Datenqualitätsschritte an die zentral aufgebauten Data Science Teams übergehen. Gleichzeitig begannen die zu verarbeitenden Datenmengen und die Geschwindigkeit mit der diese eintreffen zu steigen. Der Begriff »Big Data« prägte diese Phase. Diese neuen Aufgaben profitierten davon, dass Datenspeicher und -Verarbeitung physisch nah beieinander lagen. Gleichzeitig wurden die Serverkapazitäten hochskaliert, sodass eine massiv parallele Verarbeitung möglich war. Traditionelle Data Warehouses konnten diese neuen Anforderungen häufig nicht erfüllen, sodass sich zunehmend Data Lakes etabliert haben, welche genau diese Schwachstellen der Data Warehouses adressierten. Die ersten Data Lakes wurden dabei oft im eigenen Corporate Datacenter on-premise aufgebaut, inzwischen gibt es mit Private, Hybrid und Public Clouds attraktive Alternativen. Neben diesem technischen Infrastrukturtrend hat sich gezeigt, dass »klassische« Data Lakes auch architekturelle Schwächen aufweisen. So kommt es vor, dass die gewünschte Zuverlässigkeit der Daten und die Performance nicht erreicht werden.

1 Wikipedia, Data Warehouse, https://en.wikipedia.org/wiki/Data_warehouse

Was daher schon länger praktiziert wurde hat in den frühen 2020ern einen Namen bekommen: »Data Lakehouse« (DLH). Hierbei werden Prinzipien von Data Lakes und Data Warehouses zusammengeführt. Es gibt weiter zentrale Datenhaltung aber zumindest einige Teile davon werden wieder in einer verarbeiteten, Use Case spezifischen Sicht gespeichert: teils im Data Lake selbst, teils in ausgelagerten, spezialisierten Systemen (als Beispiel können hier Graphdatenbanken und ähnliche Systeme genannt werden).

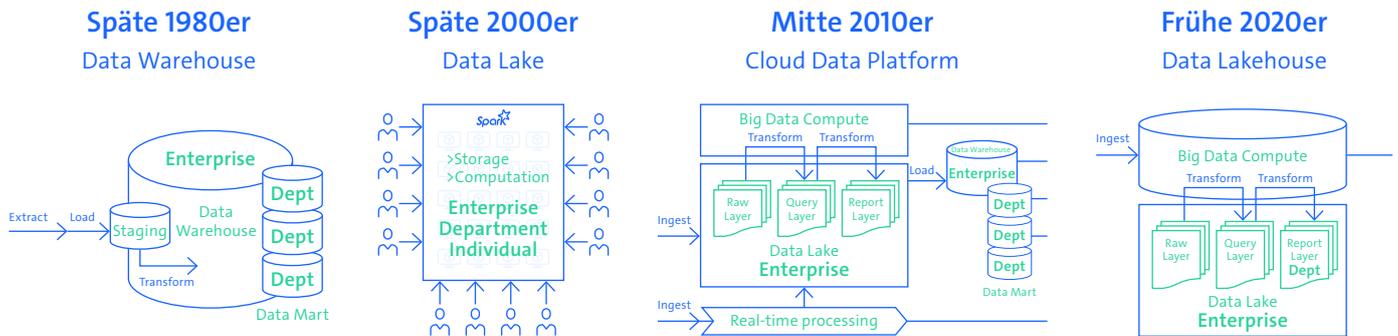


Abbildung 1: Die Evolution der Datenarchitekturen (Quelle: Nach Microsoft via Ralph Kemperdick)

Diese Konzepte verbindet also, dass sie dezentrale, operative Quellsysteme anbinden, Daten in ein zentral gemanagtes System laden, von einem zentralen Team verarbeiten lassen und dann Ergebnisse in Form von Reports oder Ergebnissen analytischer Modelle zurückgeben.

Nachteile dieser Vorgehensweisen sind jeweils die Zentralisierung der Verantwortung für die Daten im DWH, DLH oder DL bei Rollen, welche nicht für die initiale Erstellung der Daten in den operativen Systemen verantwortlich sind. Die Datenqualität gerät somit schnell aus dem Fokus. Durch die Abhängigkeit von einem zentralen Team entstehen zusätzlich Bottlenecks, die zu einer Verlangsamung der Datenbereitstellung führen.

Diese Nachteile wurden erkannt und ab 2018 von Zhamak Dehghani in einem neuen Konzept angegangen, dass sie »Data Mesh« nannte und welches Einflüsse wie Microservicearchitekturen, Domain Driven Design und DevOps aus der Softwareentwicklung auf Daten überträgt. Die große Neuheit ist hierbei nicht, dass eine neue Technologie eingeführt wird, sondern, dass die Probleme der Zentralisierung durch eine Änderung von Organisation, Data Governance und Data Culture-Maßnahmen gelöst werden sollen.

Das Data Mesh Konzept wird in der Datencommunity aktuell so interessiert diskutiert, dass es tatsächlich das nächste weitverbreitete Designpattern für Daten werden könnte.

Weitere Innovationen sind Dataspaces und Dataspace Mesh. Die Datenstrategie der Bundesregierung beschreibt einen Datenraum als »einen gemeinsamen, vertrauenswürdigen Raum für Transaktionen mit Daten. Ein Datenraum basiert beispielsweise auf gemeinsamen Standards (oder Werten, Technologien, Schnittstellen), die Transaktionen mit Daten erlauben oder befördern«².

2 Bundesregierung der Bundesrepublik Deutschland. 2021. Datenstrategie der Bundesregierung – Eine Innovationsstrategie für gesellschaftlichen Fortschritt und nachhaltiges Wachstum. Kabinettsfassung 2021-01-27, Bundeskanzleramt, Berlin, <http://www.bundesregierung.de/publikationen>

Im Kern drehen Datenräume die traditionelle Logik der Datenhaltung um. Statt alle Daten zentral zu lagern, können Daten verteilt abgelegt werden ja sogar dort verbleiben, wo sie entstehen. Denn nicht die zentrale Lagerung ist entscheidend, sondern ob eine Anwendung, wie z.B. eine Korrelationsanalyse oder ein Deep Learning Algorithmus, die richtigen Daten in ausreichender Menge bekommt. Also ein Just-in-Time Datenteilen statt zentraler Lagerhaltung. Das Problem war aber bisher, dass sich die Beteiligten einer Datentransaktion oft nicht trauen, Daten tatsächlich zu teilen. Aus unterschiedlichen Gründen, wie z.B. aus Sorge um Wettbewerbsvorteile und Datenschutz. Kurz, oft fehlt der Schutz der Datensouveränität, dem Recht, die Kontrolle über seine Daten zu behalten. Sobald eine Datei verschickt wird, kann alles mit ihr passieren. Hier kann neue Technologie helfen, so dass zwei Parteien, die sich zwar nicht trauen, aber dennoch einer Datentransaktion vertrauen können, von der sowohl ein Endkunde wie die beiden Parteien profitieren.

Mit den ersten Datenräumen entsteht auch ein Dataspace Mesh, die Verkettung und Verlinkung von Datenräumen.³ Ein Data Space Mesh oder Mesh Data Spaces erweitert den Begriff eines Data Mesh über eine Unternehmensumgebung hinaus, um neuartige verteilte, föderierte und dezentralisierte Datensysteme wie Datenräume.⁴ Merriam-Webster definiert ein Mesh als »ein netzartiges Muster« oder »eine Anordnung ineinandergreifender Verbindungen«, was besonders gut mit den ineinander verwobenen Dimensionen der Interoperabilität von Systemen korrespondiert, einschließlich technischer, semantischer, organisatorischer und rechtlicher Dimensionen.⁵

3 Drees, H., S. Pretzsch, B. Heinke, D. Wang, and C. Schlueter Langdon. 2022. Data Space Mesh and Interoperability: Example of Mobility. Technical Paper, 14th ITS European Congress (in Kürze erscheinend)

4 Siehe »7 Data Mesh Fallbeispiele« in Oracle. 2021. Enterprise Data Mesh – Solutions, use cases & case studies. Oracle & GoldenGate ebook (Fall), <https://www.oracle.com/a/ocom/docs/datamesh-ebook.pdf>

5 European Union. 2017. New European Interoperability Framework (EIF) – Promoting seamless services and data flows for European public administrations. Publications Office of the European Union, Luxembourg (2017-11-30), <https://op.europa.eu/en/publication-detail/-/publication/bca40dde-deee-11e7-9749-01aa75ed71a1/language-en>



3 Architektur

Nachdem wir gelernt haben woher das Konzept des Data Mesh kommt, wollen wir nun in die konkrete Architektur einsteigen. Der wichtigste Punkt hierbei ist, dass es sich bei dem Data Mesh um ein »soziotechnologisches Paradigma« handelt.⁶ Hinter diesem sperrigen Wort verbirgt sich der Fakt, dass es bei einem Data Mesh nicht um eine rein technische Lösung, sondern um eine Kombination von Datenstrategie, Enterprise Architektur und Unternehmensorganisation geht. Technik spielt eine unterstützende, aber untergeordnete Rolle.

Wir haben gelernt, dass das Data Mesh eine Lösung darstellt, um Bottlenecks zu umgehen und die Geschwindigkeit der Innovation innerhalb von Teams und Organisationen zu erhöhen bzw. wieder herzustellen.

Ein zentraler Baustein dieser Architektur ist daher die konsequente Identifizierung und Entfernung dieser Bottlenecks indem möglichst viel Verantwortung in verteilte Teams gegeben wird, die trotzdem anhand von zentral definierten »Leitplanken« agieren.

Dieser Gedanke mündet in einer logischen Architektur, welche auf vier Prinzipien aufbaut.⁷

⁶ Dehghani, Z. 2022. Data Mesh. O'Reilly, ↗<https://www.oreilly.com/library/view/data-mesh/9781492092384/>

⁷ Dehghani, Z. 2022. Data Mesh. O'Reilly, ↗<https://www.oreilly.com/library/view/data-mesh/9781492092384/>

3.1 Domain-oriented ownership

Die Verantwortung für Daten liegt nicht bei einem zentralen Team, sondern wird in die Hände der Teams gegeben, die am »nächsten« an den Daten arbeiten. Dies sind häufig – aber nicht immer – die Teams, die auch die Daten generieren.

Damit unterscheidet sich ein Data Mesh grundlegend von vielen Data Lake oder Data Warehouse Projekten, in denen es Teams gibt, die sämtliche Daten entgegennehmen, säubern, strukturieren und speichern, aber selbst keine Experten der jeweiligen Domänen sind, aus welchen die Daten kommen. Diese Teams generieren daher häufig qualitativ nicht optimale Datensätze und tragen zur Trägheit des Systems bei.

Das Team, welches die Daten bereitstellt, kümmert sich also um den gesamten Lebenszyklus. Nutzer sehen nur das Ergebnis – das Datenprodukt. Dies bringt uns zum nächsten Prinzip.

3.2 Data as a Product

Wenn »Zeit Geld ist«, dann ist Datenanalyse heute eine Katastrophe. In der Literatur dokumentierten Übersichten zufolge werden heute mehr als 80% des Zeitbudgets eines Datenanalyseprojekts für die Datenaufbereitung aufgewendet – nicht etwa für die Entwicklung von Algorithmen.⁸ Dies würde das 80/20 Pareto-Prinzip, einen Eckpfeiler der Geschäftseffizienz, auf den Kopf stellen.⁹ Die Lösung: Daten wie Produkte behandeln und bewährte Ansätze, wie das Produktmanagement auf Daten anwenden.¹⁰ Es geht darum, die Daten als eigenständiges Produkt zu verstehen und nicht als Nebenprodukt einer anderen Tätigkeit. Zu einem Produkt gehört alles was nötig ist dieses bereit zu stellen: 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.

In anderen Bereichen der Softwareentwicklung findet immer häufiger ein Prinzip namens »Shift-Left« Anwendung. Damit ist gemeint, dass Aufgaben und Verantwortung näher an die Quelle rücken. Die wohl prominenteste Vertreterin dieses Prinzips ist die Art und Weise wie Softwareprodukte heute erstellt werden: DevOps erhöht die Geschwindigkeit, Stabilität sowie verbessert die Abstimmungsprozesse in der Softwareentwicklung. Dies geschieht durch eine Kombination der Entwicklung (»Development«) und dem Betrieb (»Operations«) in einem einzigen Team. Eine vergleichbare Entwicklung findet sich unter dem Begriff DataOps für Prozesse rund um das Datenmanagement und Maschinelles Lernen (MLOps) sowie der Security (DevSecOps) und auch beim Testen von Software.

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

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

10 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>

Früher war es üblich, dass Entwicklungsabteilungen Produkte entwickelt und dann an Betriebsabteilungen übergeben haben. DevOps löst diese starre Verteilung auf. Das Data Mesh Prinzip dehnt diese Verschiebung der Verantwortlichkeiten nun auch auf die generierten Daten aus, welche Stand heute noch sehr häufig von zentralen Teams verwaltet werden.

Die Daten und deren Nutzbarkeit als eigenständiges Produkt werden zurückgeholt zu den Teams, die die Daten produzieren und es liegt in deren Verantwortung, ein vernünftiges Datenprodukt zu bauen.

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.

Dieses Prinzip ist eng verwandt mit dem vorherigen – Domain Ownership. Während es dort um die Zuständigkeiten und deren Dezentralisierung ging, geht es hier 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.

Ein Produkt sollte diese Anforderungen 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. Viele dieser Anforderungen können durch Technologie erreicht oder unterstützt werden, womit wir zum dritten Prinzip kommen.

3.3 Self Service Data Platform und Data Factories

Beim Lesen der vorherigen Punkte mögen Bedenken vor einer Vervielfachung von Wissen, Infrastruktur und Daten aufkommen. Auch gibt es Sorgen über den Fachkräftemangel, denn es klingt als bräuchte jedes Team genügend Fachleute, um den gesamten Lebenszyklus eines Datenproduktes abzudecken.

Das ist allerdings nicht der Fall: Ein zentrales Infrastrukturteam sollte die nötigen Tools und Standards in Form einer Plattform bereitstellen, um es nicht spezialisierten Mitarbeitenden zu ermöglichen, die notwendigen Pipelines und Verarbeitungsschritte selbst auszuführen. Dazu gehören auch Querschnittsfunktionalitäten wie Monitoring und Logging, genauso wie gemeinsame API Standards und Protokolle.

Dieser »as-a-Service« Gedanke sollte unbedingt in Kombination mit einem »as Code« Ansatz verbunden werden: Die Möglichkeit, Infrastruktur als Programmcode bereitzustellen, oft innerhalb weniger Minuten, ermöglicht eine neue Flexibilität bei der Herstellung von Datenprodukten.

Es handelt sich hier aber tatsächlich ausschließlich um die Zurverfügungstellung der Plattform sowie von Wissen. Nicht aber um eine konkrete Umsetzung – diese liegt, wie erwähnt, bei den Fachteams selbst.

Durch diese Trennung der Aufgaben ist es nicht nötig, toolspezifische Experten in jedem Team zu haben, es reichen stattdessen geschulte Anwender dieser Tools. Ziel sollte es sein, über typische SaaS/IaaS/PaaS-Angebote hinweg Data Mesh spezifische Produkte anzubieten. Hierbei kann es sich um Möglichkeiten handeln, Datenprodukte in einem zentralen Katalog zu registrieren oder die Überwachung von Datenqualität zu automatisieren.

Dadurch, dass Data Mesh noch relativ neu ist, gibt es hier noch einige Lücken in den verfügbaren Angeboten.

Einige technische und nicht-technische Themen hingegen, lassen sich aufgrund Ihrer Art nicht komplett in die einzelnen Teams auslagern, womit wir zum vierten und letzten Prinzip kommen.

Ein eleganter Lösungsansatz für die Organisation der Erstellung von Datenprodukten ist der Wechsel von Handarbeit hin zu industrialisierten Verfahren, so wie auch bei anderen Produkten und Konsumgütern – den sogenannten Data Factories (Schlueter Langdon und Sikora 2020).¹¹ Im Kern geht es um die Zergliederung des Konversionsprozesses von Rohdaten zu Datenprodukten in Kernverfahrensschritte und deren Automatisierung.

3.4 Federated Computational Governance

Es gibt einige Aufgaben, die immer zentral gesteuert werden sollten. Ein Beispiel hierfür sind die typischen Vorgaben der Rechtsabteilungen zum Thema Datenschutz oder sensiblen Daten im Allgemeinen. Auch die (Daten-)Sicherheit sowie automatisierte Tests und Monitoring gehören häufig dazu.

Dafür bietet es sich an ein föderiertes Team aufzubauen, welches aus Mitgliedern aller Abteilungen besteht (Domänen-/Fachabteilungen, Rechtsabteilung, Sicherheit usw.). Dieses Team ist dafür zuständig, zentrale Regeln zu definieren und ggf. Vorgaben für Automatismen zu machen.

Die einzelnen Produktteams hingegen sind für ihre lokale Governance verantwortlich, welches z.B. auch Autorisierungsregeln auf die Datenprodukte beinhaltet.

Um diese beiden Komponenten – die lokale und globale Governance – zu verzahnen, bietet es sich an, möglichst viel auf »as code« Ansätze zu bauen. In diesem Falle also »Policy as Code« oder »Security as Code«.

¹¹ 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, Schweiz: 43-55, <https://www.springerprofessional.de/creating-a-data-factory-for-data-products/18815012>

3.5 Zusammenfassung

Diese vier Prinzipien sind dafür gedacht, zusammen genutzt zu werden – einzeln umgesetzt sind sie nicht ausreichend, um die Ziele zu erfüllen. Jedoch komplementieren sie sich. Zusammengenommen sind diese vier Prinzipien hingegen hinreichend.

Eine umfassende Landschaft zur Verarbeitung von analytischen Daten umfasst mehrere Komponenten, welche von unterschiedlichen Rollen in der Organisation wahrgenommen werden.

Die logische Architektur eines Data Mesh stellt sich beispielhaft wie unten abgebildet dar:

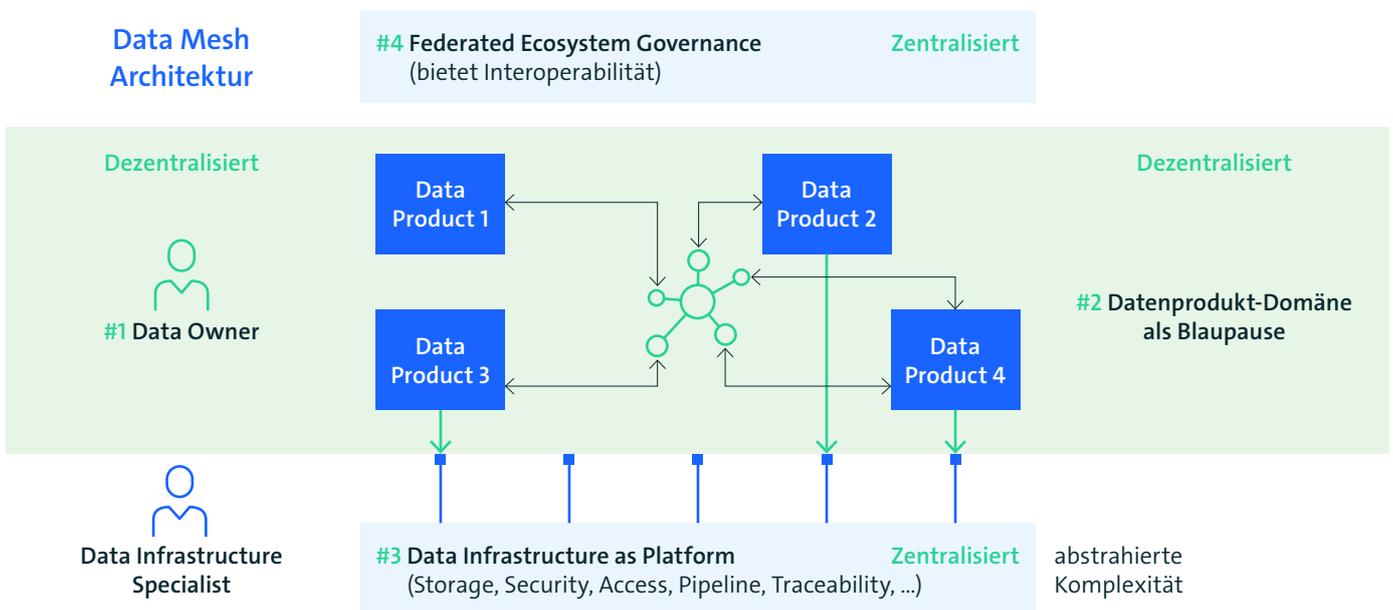


Abbildung 2: Logische Architektur (Quelle: Nach Microsoft via Ralph Kemperdick)

Zusammen umgesetzt, erlauben diese vier Prinzipien eine dezentrale Verarbeitung und Zurverfügungstellung von Daten, was es erlaubt, schneller auf Änderungen oder neue Anforderungen zu reagieren.

Ein Data Mesh umzusetzen, bedeutet Änderungen in der Organisation zu erlauben und zu implementieren. Es kann zur Auflösung vorhandener Teams kommen, zu neuen Aufgaben in vorhandenen Teams, Budgetfragen müssen geklärt werden und vieles mehr. Auch darf nicht vergessen werden, dass Data Meshes noch sehr neu sind (2018) und es noch einige ungeklärte Fragen und technische Lücken gibt.

In der Tat werden mit der Erzeugung von Datenprodukten und deren Bereitsstellung sowie Katalogisierung zum schnellen Auffinden/Verstehen Datensilos verhindert. Nichtsdestotrotz sind auch Datenprodukte ihrerseits domain-spezifisch und lassen häufig keine Verbindung mit anderen Datenprodukten zu. Dennoch hebt erst die Kombination von mehreren Datenprodukten das eigentliche Potential, dass in den Daten liegt. Das Datenprodukte ihrerseits Silos darstellen können liegt daran, dass die

Datenprodukt-Owner oft keine domain-spezifischen Kenntnisse über die anderen Datenprodukte besitzen. Folglich muss es ein zentrales (Cross-)Team geben, das nunmehr Kenntnis von mehreren Datenprodukten besitzt und Verbindungen/Überführungen durch das Erzeugen von neuen Kombi-Datenprodukten herstellt. Eine andere Option wäre es, dass man ein übergeordnetes Domain-Modell zwischen den Datenprodukten schafft, das eine Überführung/Austausch von Daten aus unterschiedlichen Domänen schafft, wozu wieder ein zentrales Cross-Team erforderlich ist. Die Dezentralisierung des Data Mesh hat also seine Grenzen, wenn es um die Definition übergeordneter Gemeinsamkeiten/Mapping der Datenprodukte geht, welche bei der Implementierung eines Data Mesh bedacht werden müssen.

4 Die Bedeutung für Unternehmen

Während die Datenmengen (»data assets«) und damit deren wertstiftendes Potenzial in Unternehmen tendenziell wachsen, hinkt die tatsächliche gewinnbringende Datennutzung häufig hinterher. Aus den vorangegangenen Kapiteln wird ersichtlich, dass Daten erst nutzbar werden und gewinnbringend eingesetzt werden können, wenn diese eine hohe Qualität aufweisen, zugänglich und aufbereitet sind und von den potenziellen Nutzern verstanden werden.

Entscheidend dabei ist, dass ein nachhaltiges operatives Modell gefunden werden muss, welches effizient und effektiv Nutzen erbringt – denn keine Organisation sammelt und analysiert Daten zum reinen Selbstzweck oder Spaß.

Durch die Anforderung, Daten im wachsenden Ausmaß zu nutzen, entsteht hier ein neues Tätigkeitsfeld im Unternehmen, welches an der Schnittstelle zwischen IT und Datenanwendern (Data Scientists, Data Analysts, etc.) liegt: Die aktive, intelligente und nutzerorientierte Datenhaltung und -aufbereitung. Diese Rolle ist geprägt durch die Verantwortung für die Haltung und Entwicklung »nutzbarer Daten«. Dieses Tätigkeitsfeld in eigenen, zentralen Rollen für die einzelnen »Datenpools« zu professionalisieren ist aus mehreren Gründen eine vielversprechende Vorgehensweise:

Aufbereitete Daten erleichtern den Datenanwendern die Arbeit. Diese kommen schneller zum Ziel, einen **Nutzen aus Daten** zu generieren. Das zahlt auf einen positiven »return on data invest« ein, denn:

1. Eine gesteuerte Datenaufbereitung mit klaren Verantwortlichkeiten sorgt für einheitliche Definitionen und der Vermeidung von Fehlern bei der Datenvorverarbeitung und trägt somit zur **Qualität der Daten** bei. Zudem wirkt sie der Doppel- bzw. Mehrfacharbeit durch Datenanwender an verschiedenen Stellen eines Unternehmens entgegen.
2. Die Daten werden im Sinne der Demokratisierung nicht nur für hochgebildete Datenspezialisten sondern auch eine **breitere Menge an Mitarbeitenden** nutzbar. Das Konzept wirkt der Entwicklung und Verfestigung von Silos entgegen und verbreitert tendenziell das Spektrum, in dem die Daten für ein Unternehmen genutzt werden können.
3. Durch die neuen, **klaren Verantwortlichkeiten** für aufbereitete Daten, die Datenqualität und den breiteren Nutzerkreis, ist mit einem positiven Kosten-Nutzen-Effekt zu rechnen.

Dass dieses Aufgabenfeld als ein laufendes, langfristiges Tätigkeitsfeld angelegt wird, ist vor dem Hintergrund der laufenden Dynamik von Datensätzen/Data Warehouses/Data Lakes einerseits sowie der sich wandelnden Anforderungen der Datennutzer

andererseits wichtig. Stellt man sich die Unternehmensdaten als Weizen auf einem Feld vor, gäbe es da zunächst eine Person (z.B. Business Analyst), welche die Saat austrägt (generierte Daten in operativen Systemen werden in die Datenspeicherung eingebracht), und einige Monate später andere Personen, welche auf dem Feld ein bestimmtes Produkt ernten (z.B. Data Engineers) und wieder andere, die die Ernte dann weiterverarbeiten möchten (Data Analysts, Data Scientists etc.) oder im Feld nach entsprechender »Auswahl« neu aussähen. Das Konzept von Data Mesh und Datenprodukten sorgt dafür, dass das Feld in der Zwischenzeit gepflegt wird, denn nur so kann die Ernte eine hohe Qualität aufweisen. Die Lücke zwischen der alleinigen Haltung von Rohdaten in anwenderfernen Teams einerseits und der Verarbeitung von aufbereiteten Daten in Data Science und Analytics Teams andererseits wird folglich geschlossen, in dem für die nutzerorientierte Datenaufbereitung und -haltung ein eigener Verantwortungsbereich geschaffen wird, der sich in vielfältiger Weise für ein Unternehmen lohnt.

Ein entscheidender Aspekt ist hierbei die Fähigkeit zur Skalierung über eine ganze Organisation hinweg: Während in einer initialen Phase eine zentralisierte Organisationsform zur Verwaltung und Speicherung von Daten durchaus Sinn ergeben kann, ist bei wachsender Menge von Nutzern (bzw. Datenprodukten) schnell eine Situation erreicht, in der die Daten sich von den Erzeugern und Nutzern »entfremden«. In diesem Fall sind die Weiterverarbeitung, Speicherung und Nutzung in einer einzigen (monolithischen!) Plattform eher hinderlich als förderlich. Durch Proliferation der Daten hin zu mehr Datenquellen und komplexeren, potentiell unstrukturierten Datensätzen wird eine Beschreibung, Katalogisierung oder gar Versionierung von zentraler Stelle aus zu einer Sisyphusarbeit oder gar Bottleneck. Das inhärente Verständnis der Datensätze wird zudem erfahrungsgemäß schwieriger, je weiter man sich von der ursprünglichen Quelle entfernt, was die erfolgreiche Nutzung zusätzlich erschwert, und gerade in multinationalen Organisationen oftmals zur Nicht-Nutzung außerhalb des Ursprungssilos führt. Gerade durch die Anforderung zu schnelleren Entwicklungszeiten und durchgreifender Innovation kann eine »Skalierung in die Breite« als Alternative zum Aufstocken einer »Zentraleinheit« gesehen werden – gerade, wenn es mit Legacy-Systemen noch viele »Baustellen« gibt die aufgeräumt werden sollten.

Gerade im Umfeld großer Unternehmen ist ein Blick auf den längeren Lebenszyklus durchaus hilfreich – denn Akteure und Zuständigkeiten wechseln, und relevante Detailinformationen zur »Historie« und den Details gehen dabei oft verloren. Beispiele hierfür sind z.B. Transaktionsdaten aus ERP-Systemen, die in Beschreibung der Struktur nicht über Feldbezeichnungen wie »Datum« hinausgehen: Doch um welches Datum handelt es sich? Exakt an einem solchen Punkt wird auch die Tragweite der Herausforderung klar. Weiterhin sind die Technischen Experten oftmals weit vom eigentlichen Geschäft (und somit der wertstiftenden Nutzung) entfernt. Auch gilt, je größer die Organisation wird, desto größer wird diese Herausforderung.

5 Organisation, Führung und Kultur im Data Mesh Kontext

Data Mesh basiert auf vier zentralen Konzepten, von denen das vierte die Federated Computational Governance ist. Beim Ansatz des Data Mesh geht es dabei um den Wechsel von zentralisierten Datenteams und -architekturen zu einem dezentralen Gesamtmodell. Passend dazu muss auch geklärt werden, welche Governance-Funktionen weiterhin in den Zuständigkeitsbereich der zentralen IT-Organisation fallen und was nun in der Verantwortung der jeweiligen Domäne liegt. Während die allgemeine Verlagerung bei Data Mesh von einer zentralen Architektur zu einer dezentralen Architektur erfolgt, gibt es Aspekte der Data Governance, die notwendigerweise bei einer zentralen Gruppe verbleiben müssen, damit ein einheitlicher Ansatz verfolgt und sichergestellt ist, dass alle Aspekte der Governance angemessen berücksichtigt werden.

Das Data Mesh schlägt eine geteilte Verantwortung zwischen den Domänen und der zentralen IT-Organisation vor, um die Governance einzuhalten und gleichzeitig eine angemessene Autonomie für die Domänen zu ermöglichen. Es ist dabei wichtig zu beachten, dass es sich um ein föderiertes Modell handelt, d.h. es gibt eine domänenübergreifende Vereinbarung darüber, welche Aspekte der Governance an zentraler Stelle auf der Mesh-Ebene und welche dezentral von den Domänen gehandhabt werden.

5.1 Data Governance

Governance kann viele Dinge umfassen, vor allem aber geht es beim Data Mesh darum, 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 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 Benutzer zugänglich?
- **Qualität:** Wie wird die Datenqualität in passender Weise quantifiziert und den Nutzern mitgeteilt?
- **Entitätsstandardisierung:** Gibt es eine Einigung zur Terminologie in verschiedenen Bereichen?
- **Provenance:** Ist klar, wer für die Daten verantwortlich ist und woher sie stammen?

Es liegt deshalb auf der Hand, dass einige dieser Governance-Aspekte auf der Mesh-Ebene festgelegt werden, während andere im Ermessen der jeweiligen Bereiche liegen. Data

Mesh ermöglicht diese geteilte Verantwortung durch ein Federated Governance-Modell, das globale Standards und Richtlinien ermöglicht und domänenübergreifend angewendet werden kann. Dabei wird den Domänen ein hohes Maß an Autonomie eingeräumt, wie sie diese Standards und Richtlinien innerhalb ihrer eigenen Domäne umsetzen. Da die Governance im Data Mesh auf ein Gleichgewicht zwischen Interoperabilität und globalen Standards setzt, 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- und Infrastructure-as-Code-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 domänenübergreifende Zusammenarbeit voraus, eine weitere sehr große Herausforderung in großen Unternehmen mit ihren vielfältigen Interessendivergenzen. Ein effektiver Weg, dieser Herausforderung zu begegnen, ist es, jeweils einen Vertreter aus jeder Domäne, als Mitglied in einen domänenübergreifenden Governance-Council zu entsenden, der hier mit der zentralen IT- und Governance-Organisation zusammenarbeitet und die zwingend notwendige Kommunikation unter- und miteinander unterstützt.

5.2 Data Stewardship

Wie in den Abschnitten zuvor hinreichend belegt, bringt die Arbeit mit und die Organisation rund um Daten vielerlei Probleme mit sich. So weisen *Daten* in der Regel die Eigenschaft auf, dass sie sich *nicht selbst erklären*. An einer Stelle muss immer eine Interpretation von Daten geliefert, die richtige Verwendung definiert oder über eine Qualität entschieden werden, dies geschieht üblicherweise durch Metadaten. Zusätzlich dazu sind Daten in der Lage, durch verschiedene Geschäftsbereiche zu wandern. Anders als ein physisches Gut können Daten dabei auch zeitgleich an verschiedenen Stellen genutzt werden. Diese und weitere Punkte führen dazu, dass eine inhärente Verantwortlichkeit für Daten und Datenthemen schwierig zu bestimmen ist.¹² Data Stewardship definiert genau diese Verantwortlichkeit und kann dementsprechend als angewandte Data Governance verstanden werden. Das Ziel der Data Stewardship liegt dabei in der klaren Zuweisung von Verantwortlichkeiten und dem verstärkten Commitment der Stakeholder im Datenmanagement. Im Kontext des Data Mesh sorgt also die Stewardship für den organisatorischen Rahmen, um einen Domain Driven Ansatz überhaupt angehen zu können. Erfolgsfaktor für eine erfolgreiche Data Stewardship ist das Bewusstsein über die technische Zentralisierung der Dateninfrastruktur und die organisatorische föderativ zugewiesene Verantwortung.

¹² Plotkin, D. 2014. Data Stewardship: An Actionable Guide to Effective Data Management and Data Governance, <https://www.sciencedirect.com/book/9780124103894/data-stewardship>

Somit setzt man eben weder auf eine reine technische noch eine reine fachliche Zuordnung. Vielmehr setzt man auf eine geteilte Verantwortlichkeit, die sich sowohl an fachlichen Use Cases als auch an technischen Abhängigkeiten orientiert.

Zu den spezifischen Rollen findet man in der Literatur vielfältige Ausprägungen und Varianten dabei sind folgenden Rollen immer wesentlicher Teil der Rollenkonzepte:

- Data Owner
- Data Steward

Der Data Owner trägt dabei die gesamte Verantwortung für einen fachlichen Datenbereich, während der Data Steward für die Operationalisierung einer Bereichsstrategie, die Erhebung und Kontrolle der Datenqualität sowie das Erteilen von Berechtigungen für den Datenzugriff verantwortlich ist.¹³ Hierdurch gelingt es, Verantwortlichkeiten an bereits etablierte Berichtslinien anzupassen und sowohl in der Breite als auch in der Tiefe der Organisation zu verankern. Um dennoch eine übergreifende Instanz einzurichten werden sogenannte Councils installiert. In diesen Councils findet ein horizontaler Austausch statt und das Kollektiv aus Data Ownern bzw. Data Stewards stellt somit weitestgehend autonom eine generelle Standardisierung und Interoperabilität innerhalb der Organisation sicher.

5.3 Data Culture

»Culture eats strategy for breakfast«

Dieses Zitat von Peter Drucker ist für die Implementierung des Data Mesh aktueller denn je. Für den Erfolg des Data Mesh spielt die Kultur eine besondere Rolle, denn sie prägt die Art und Weise wie menschliche und technische Entitäten interagieren, wer Zugang zu Daten hat und unter welchen Bedingungen dieser stattfindet. In Bezug auf eine für das Data Mesh förderliche Kultur des »Data Sharing is Caring« lassen sich fünf wesentliche Punkte herausstellen:

- Verantwortungsbereit
- Offen und vertrauensvoll
- Wissbegierig fragend
- Iterativ lernend
- Visionsgetrieben

Verantwortungsbereit

Da Daten laut Data Mesh Konzept zu nutzbaren Datenprodukten weiterentwickelt werden, entstehen neue Verantwortungsbereiche allein für die Entwicklung und Bereitstellung von Daten. Daten sollen in einem angemessenen Zeitrahmen einem breiten Anwenderkreis in nutzbarer Form zur Verfügung gestellt werden. Ist Verantwortungsbewusstsein und Streben nach qualitativ hochwertigen Produkten im Unternehmen verankert, erleichtert dies die Einführung von Data Mesh Strukturen.

Offen und vertrauensvoll

Daten sollen im Sinne des Unternehmens im Rahmen der gesetzlichen Regeln in vielfältiger Weise genutzt werden können. Von Vorteil ist eine Kultur, in der Mitarbeitende gemeinsam, team- und fachübergreifend auf Unternehmensziele hinarbeiten.

Wissbegierig fragend

Daten sind für andere häufig nicht selbsterklärend. Um nutzbare und qualitativ hochwertige Datenprodukte zu entwickeln, sind zwei Tätigkeiten in enger Abstimmung mit allen Beteiligten besonders wichtig:

- Datenverarbeitungsschritte laufend hinterfragen und reflektieren und
- Konsens zu Definitionen von Datenfeldern finden.

Zudem ist eine Offenheit in Bezug auf neue Technologien gefragt, um Prozesse rund um Daten laufend optimieren zu können.

Iterativ lernend

Eine Organisation, welche agil mit Daten und Technologie umgeht, kann schnell die für sich passenden Methoden und Strukturen rund um Datenprodukte finden. Hierbei helfen agile Frameworks wie Scrum, Kanban und Design Thinking. Bei »Fail Fast« liegt der Fokus auf dem Lernen aus der Vergangenheit als Input für die Zukunft, und nicht in der Schuldzuweisung vergangener Ereignisse.

Visionsgetrieben

Organisationen, welche Daten als ein neues Entwicklungsfeld im Unternehmen sehen, sehen sich eventuell durch längere Zeit der Nicht-Investition in technische Infrastrukturen mit einer hohen »technical debt« und somit großen Hürden der unternehmerisch wirksamen Datennutzung ausgesetzt. Das Konzept Data Mesh kann dabei als Angebot gesehen werden, Daten und Technologien Schritt für Schritt zu entwickeln, um langfristig eine größere Vision bzw. »Data Strategy« umzusetzen. Zentral ist dabei, Daten- und Technologiestrategien fest in der Unternehmensvision zu verankern.



6

Einsatzbeispiele

Aufbauend auf den nun etablierten Anwendungsgrundlagen gehen wir nun auf konkrete Praxisbeispiele ein, welche einen konkreten Eindruck zu Data Mesh Initiativen liefern sollen. Data Mesh wird bereits in vielen Unternehmen und Organisationen auch im deutschsprachigen Raum genutzt, wie folgende Praxisbeispiele überzeugend nachweisen. Die vorgestellten Praxisbeispiele sind geeignet, Lesende zu inspirieren, in ihren Unternehmen Einsatzmöglichkeiten und spezifische Vorteile für Data Mesh zu entdecken und damit verbundene Daten- und Geschäftsmehrwerte zu heben.

6.1 Einsatzbeispiel Allianz Direct

Data Mesh bei der Allianz Direct – Wie eine partizipative Datenkultur, moderne Technologien und integrative Prozesse schnell Datenmehrwerte realisieren

Steckbrief

Die Allianz Direct ist der Online-Direktversicherer der Allianz und versteht sich als Technologiefirma mit einer Versicherungslizenz. Kundenzentriertheit, Einfachheit und Direktheit stellen die Erfolgsfaktoren der Allianz Direct dar. Dazu werden vermehrt digitale Technologien in der Kundeninteraktion, zur Personalisierung der Versicherungs-journey und für das Treffen schnellerer und datengetriebener Entscheidungen genutzt.

Ausgangslage

Die Allianz Direct nutzt Daten in verschiedensten Anwendungsfällen im Versicherungskontext z.B. zur Personalisierung von Versicherungsprodukten. In der Vergangenheit wurde die Arbeit mit Daten hauptsächlich von zentralen Teams durchgeführt. Die Konfrontation mit wachsenden Anforderungen und unzureichende technische Gegebenheiten bremsen jedoch die schnelle und erfolgreiche Umsetzung von Datenprojekten stark aus.

Herausforderungen

Die bisherige Organisation war geprägt durch das zentrale Business Intelligence Team, als technischer Partner für die Fachbereiche. Die stetig wachsende Zahl an Datenprojekten, die zudem nur mit Fachbereichswissen erfolgreich sein können, führten schnell zu Ressourcenknappheiten oder zu langen Projektlaufzeiten. Weitere Reibungsverluste führten zu einer nicht vollständig und in aller Tiefe umgesetzten Data Governance, fragmentierten Technologieplattformen und Datenbeständen.

Für einen effektiveren und effizienteren Umgang mit Daten waren somit nicht nur eine, bereits angestoßene, technische Transformation der Daten und IT-Systeme notwendig, sondern insbesondere ein kulturelles Umdenken. Hierzu sollte Data Mesh als Paradigma eine tiefere organisationale und kulturelle Durchdringung erwirken, um der Allianz Direct eine ganzheitliche Basis für bestehende, aber auch zukünftige, Datenprojekte bereitzustellen.

Lösung

Die Data Mesh Transformation fokussierte drei Kernbereiche: Erstens sollten **Mitarbeiter und die Organisation** in funktionsübergreifenden Teams, bestehend aus Domänen- und Datenexperten, dezentral und lösungsorientiert zusammengeführt werden. Die heterogene Teamzusammenstellung ermöglicht eine frühe Einbindung der Produkt- und Datenteams in Entwicklungsarbeiten. Gleichzeitig konnten so Verantwortlichkeiten klarer definiert und sinnstiftende Interaktionen gefördert werden. Zusätzlich führt

Kontakt

Vaibhavi Rege

Central Product Owner

Allianz Direct

Vaibhavi.rege@allianzdirect.nl

Martin Günther

Senior Consultant, AI & Data

Deloitte Consulting

mguenther@deloitte.de

diese kontinuierliche Kooperation und technische Unterstützung zu einer nachhaltigen Befähigung der Domänenexperten. Auch wenn die technische Infrastruktur weiterhin zentral betreut wird, können Anwender somit auch ohne technisches Know-how eigenständig mit ihren Daten arbeiten und Mehrwerte aus diesen generieren.

Zweitens stellten neue **Prozesse** einen wichtigen Erfolgsfaktor bei der Umsetzung des Data Meshs dar. Beispielsweise wurden vereinfachte Datenanbindungs- und Entwicklungsprozesse etabliert und so auch für nicht-technische Nutzer zugänglich gemacht. Allgemein stellt Self Servicing nun einen wesentlichen Bestandteil im Umgang mit den Datenressourcen dar. Das beschleunigt die Wertgenerierung, fördert so die Datenkultur und entlastet die technischen Experten.

Nicht zuletzt wurden neue **Technologien** etabliert die Self Servicing durch vereinfachte Nutzerführung, komplexitätsreduzierte Datenorganisation und automatisiertes Daten-Mapping fördern.

Mehrwert

Die Einführung von Data Mesh bei Allianz Direct hat weitreichende Mehrwerte geschaffen. So hat beispielsweise die Bildung von funktionsübergreifenden Teams, aus Domänen- und Datenexperten, zu einer verbesserten und beschleunigten Kommunikation und einer modernen Datenkultur geführt. Diese Datenkultur hat insbesondere dazu beigetragen, dass Daten als eigenständige Produkte gesehen werden die neben bess. Die frühzeitige Unterstützung durch BI- und Data-Engineering-Spezialisten hat den Domänenexperten geholfen Hürden im Umgang mit Daten zu überwinden und Eigenständigkeit zu fördern. Umgekehrt wurden die Datenexperten entlastet, da Arbeiten zunehmend von den Konsumenten in den Fachbereichen selbst durchgeführt werden können. Verstärkt wurde dies durch technische Errungenschaften wie z.B. durch Streaming und der damit verbundenen Reduktion der Komplexität.

Empfehlungen

- In großen Dimensionen denken, aber klein anfangen, mit Fokus auf klare Anwendungsfälle und einem ganzheitlichen Ansatz
- Skalierbarkeit forcieren, um eine solide Grundlage für ein nachhaltiges Wachstum zu schaffen
- Kontinuierlich in die Fähigkeiten des Teams investieren, um sie für zukünftiges Self Servicing und dafür notwendige Tools zu befähigen
- Produktteams frühzeitig einbinden, um sie auf die Data Mesh-Reise mitzunehmen
- Auch auf Management-Level ein Umdenken bewirken, um auf allen Hierarchieebenen Durchdringung der neuen Prozesse, Organisation und Kultur zu bewirken und nachhaltige Unterstützung zu erlangen
- Data Mesh erfordert nicht nur einen Wandel in der Technologie und Architektur, sondern auch in der Denkweise

6.2 Einsatzbeispiel DATAbility

Verkehrswende einfach machen?! Das Data Mesh als Enabler für die Mobilität der Zukunft.

Steckbrief

DATAbility GmbH | Darmstadt (bitkom Startup @ Get Started)

- Entscheidungsunterstützungslösungen zum Asset Monitoring & Asset Management
- Softwareprodukte & Dienstleistungen, 6 Angestellte, ca. 0,5 Mio. Euro Umsatz

vedisys AG | Griesheim

- Software für Verkehrsunternehmen & ÖPNV
- Softwareprodukte & Dienstleistungen, 15 Angestellte, ca. 1,5 Mio. Euro Umsatz

Ausgangslage

Heutige Verkehrssysteme sind von einer enormen Heterogenität geprägt, sind wenig miteinander vernetzt sowie überwiegend regional und dezentral verwaltet. Eine übergeordnete Informationszusammenführung oder Anreicherung erfolgt nicht, sodass die notwendige datentechnische Vernetzung zwischen relevanten Stakeholdern wie Verkehrsverbünde, Verkehrsunternehmen, Mobilitätsanbieter, Service-, Wartungs- und Instandhaltungsdienstleister sowie Mobilitätsnutzende bis dato ausbleibt.

Herausforderungen

Die Mobilitätswende stellt ein Schlüsselement zur Steigerung der Nachhaltigkeit von Verkehr auf Natur und Umwelt dar. Für deren Umsetzung sind neben der Mobilisierung der Gesellschaft ebenso die Steigerung der technischen Verfügbarkeit sowie die Zusammenführung relevanter Stakeholder auf Angebots- und Nachfrageseite essentiell. Aufgrund der Vielzahl proprietärer Systeme sowie dem vorherrschenden Digitalisierungsbedarf der Verkehrsträger ist dies ohne ein Paradigmenwechsel für die Anbindung und Vernetzung bestehender Daten- und Informationslandschaften jedoch nicht möglich. Damit sind der betrieblichen Optimierung von Ressourcen wie Trassen, Fahrzeuge und Personal sowie Prozessen, Services und Angebote Grenzen gesetzt – das hemmt Möglichkeiten, neue Datenprodukte zu entwickeln und anzubieten.

Lösung

Ein Data Mesh fungiert als Basis für die Zusammenführung und Verknüpfung von Daten und Informationen für die Mobilität der Zukunft: Im sogenannten mobility portal wird Angebot und Nachfrage bedarfsgerecht konsolidiert, um multimodale Mobilitätsangebote an einem zentralen Ort anbieten und vermarkten zu können sowie buchbar zu machen. Ermöglicht wird dies durch intelligente Werkzeuge (KI), sodass die Verarbeitung und Nutzung zusammengeführter Merkmale und Quellen nun eine wertschöpfende Entwicklung für neue Datenprodukte und Services ohne Medienbrüche erlaubt – bspw. für Bedarfsprognosen. Das Data Mesh setzt die Merkmale unterschiedlicher Daten- und Informationsquellen in Beziehung. Übergeordnet dienen Mediatoren für die effiziente Abwicklung von Anfragen sowie zur Konsistenzhaltung

Kontakt

Sebastian Baumann

Gründer & Geschäftsführer

DATAbility

baumann@datability.ai

der Daten. Die Anbindung zusätzlicher Datenanbieter, bspw. durch Nachrüstung bestehender Fahrzeuge mit Telematik (IoT) oder die Einbeziehung von Umweltdaten (Wetter oder Social Media) ist ebenfalls möglich. Nicht zuletzt lassen sich Services unternehmensübergreifend aufbauen, synchron halten und weiterentwickeln. Funktionen im mobility portal lassen sich auf die Bedarfe der Datenkonsumenten individualisieren, um bspw. Kennzahlen quantifizieren, Betriebsleistungen bewerten, Trends und Anomalien analysieren, Abläufe effizienter und kostenoptimal planen sowie Maßnahmen frühzeitig ableiten zu können. Im Ergebnis können effizientere Entscheidungen getroffen, maßgeschneiderte Kundenerfahrungen geschaffen und Handlungsempfehlungen an Akteure ausgeleitet werden.

Mehrwert

Ein Data Mesh schafft neue Möglichkeiten für Services und Datenprodukte, sowohl zur Reformation des derzeitigen ÖPNV-Betriebes als auch zum Aufbau zukünftiger multimodaler Individualverkehre. Für Betreiber wird dadurch einerseits ein umfassenderes und dynamischeres Asset Management & Monitoring unterstützt. Andererseits werden auf Anbieterseite Kundenbeziehungen gestärkt, indem nutzer- und serviceorientierte Aspekte stärker in den Fokus gestellt werden können. Datenkonsumenten können bedarfsgerecht mit angeforderten Merkmalen versorgt, über Änderungen oder Neuerungen im Datenraum sowie bei Services informiert und in die Lage der selbstständigen Nutzung des Datenraumes versetzt werden. Damit lassen sich Verwaltungsredundanzen insbesondere für Stammdaten sowie ineffiziente Medien- und Informationsbrüche vermeiden. Im Gegensatz zu konventionellen Architekturen können damit effizientere Zugriffs- und Bereitstellungszeiten erreicht werden.

Empfehlungen

Trotz aller Vorteile eines Data Mesh, die insbesondere individuell und domänenspezifisch ausfallen, für den genannten Anwendungsfall, sind Aufwände für Aufbau und Pflege nicht außer Acht zu lassen. Insbesondere Mapping, Synchronisierung und Aktualisierung des Data Mesh erfordern eine fortlaufend intensive Zusammenarbeit aller Datenkonsumenten (Product Owner) und Datenanbieter (Data/Domain Owner). Zudem sind Beziehungen unter Merkmalen bedarfsgerecht aufzubereiten und mit weiteren Meta-Informationen anzureichern, um die Potentiale der engen Verzahnung von Quellen und Produkten/Services nutzen zu können. Dies ist nicht nur initial, sondern vielmehr rollierend zu beachten. Insbesondere Datenkonsumenten sollten laufend Anforderungen sowie Ergebnisse aus der praxisnahen Verwendung an Datenanbieter zurückspielen. Nicht zuletzt sollte das Prinzip der Sparsamkeit (von Big zu Smart Data) besondere Beachtung finden, um Effizienzeinbußen eines Data Mesh aufgrund von zunehmender Komplexität und Größe frühzeitig zu entgegen – Aussichten auf Erfolgsgarantien gibt es nicht.



6.3 Einsatzbeispiel Deutsche Bahn

Data Mesh-Transformation durch umsetzungsorientierten Ansatz nachhaltig initiiert

Steckbrief

Deutsche Bahn AG

- Data Intelligence Center im Vorstandsressort Digitalisierung und Technik des DB-Konzerns
- Der Deutsche Bahn Konzern ist ein führender Mobilitäts- und Logistikanbieter mit klarem Fokus auf Schienenverkehr in Deutschland.

Ausgangslage

Daten werden im DB-Konzern jeweils lokal im Fachbereich oder pro Geschäftsfeld (z.B. DB Netz, DB Fernverkehr oder DB Regio) in zentral aufgebauten Analytics-Plattformen in Form von Data Warehouses oder Data Lakes verarbeitet. Direkter Datenaustausch findet heute zwischen Geschäftsfeldern und innerhalb dieser statt, ist nicht standardisiert und wird mit hohen Vorlaufzeiten im Manufaktur-Modus entwickelt.

Herausforderungen

Der Zugang zu analytischen Daten scheitert zurzeit häufig an der Verfahrens- oder internen Unternehmensgrenze und blockiert so die Weiterentwicklung zu einem über Daten integrierten Unternehmen. Die einzelnen Analytics-Plattformen entwickeln sich jeweils unabhängig voneinander auf Geschäftsfeldebene, ohne Synergien zu heben. Gleichzeitig sind Daten in diesen Systemen nicht auf eine darüberhinausgehende Nutzung ausgerichtet.

Data Mesh schien als ein dezentrales Konzept mit konsequentem Enablement in der Fläche gut zu Struktur und Kultur der Deutschen Bahn zu passen. Als nächster Schritt sollten die Rahmenbedingungen genauer untersucht werden, um abzuleiten, wie eine Transformation in Richtung Data Mesh einzuleiten ist.

Lösung

Die Deutsche Bahn verfolgt einen umsetzungsorientierten Ansatz bei der Data Mesh-Transformation, bei dem aus konkreten Pilotimplementierungen Rückschlüsse für die Transformation des gesamten Konzerns gezogen werden sollen. Im Rahmen eines konkreten Data Use Case wurden dazu die Data Mesh-Prinzipien beobachtet, um Leitplanken für die weitere Einführung des Konzepts zu entwickeln.

Der Use Case sollte dazu bewusst von einem Team umgesetzt werden, das bereits extern Erfahrungen mit Data Mesh gemacht hatte. Auf diese Weise konnten die ersten Pain Points identifiziert werden und anhand konkreter Projekterfahrung auch der Nutzen des Data Mesh-Ansatzes in die Organisation hinein kommuniziert werden.

Kontakt

Carsten Fritsch

Lead Architect Data Intelligence
House of Data – Data Intelligence Center
Deutsche Bahn
Carsten.C.Fritsch@deutschebahn.com

In der Umsetzung ergaben sich vor allem drei große Herausforderungen:

- **Ownership:** Die Zuständigkeit und Verantwortung über die dem Datenprodukt zugrundeliegende IT-Infrastruktur stellte sich in der auf effiziente Ressourcennutzung und stabilen Betrieb ausgerichteten IT-Landschaft als herausfordernd heraus. Kurzfristig konnte durch Aufbau eigener Infrastruktur auf einem Kubernetes-Cluster im Use Case die entsprechende Autonomie zurückgewonnen werden.
- **Deployment Pipelines:** Diese waren auf die Trennung von Entwicklung und Betrieb ausgerichtet, wodurch sie nicht selten durch mehrere Teams führten. Dies konnte als Ursache für langwierige Entwicklungs-Iterationen identifiziert werden. Vorläufig konnten die Probleme durch eine verbesserte Kommunikation zwischen Entwicklungs- und Betriebsteam abgemildert werden.
- **Fehlende Self Service-Dateninfrastruktur:** Der Aufbau eines Service, der die selbstständige Abfrage von Daten bei Bedarf ermöglicht, wurde als notwendig angesehen, um die angestrebte Datentransparenz sinnvoll nutzen zu können. Die Implementierung einer entsprechenden Applikation wurde angegangen und wird sukzessive erweitert.

Mehrwert

Die klar domänenorientierte Architektur ermöglicht es dem Entwicklungsteam, sich auf eine Geschäftsdomäne zu konzentrieren und dort direkten Mehrwert zu schaffen. Durch das Prinzip, Daten als Produkt zu betrachten, wurde es wichtig, sich mit der Perspektive ihrer Nutzer:innen intensiv auseinanderzusetzen. Dazu arbeitete das Team gleich zu Projektbeginn an einer Produktvision und messbaren Hypothesen. Die gemeinsame Zielrichtung war so für Entwicklerteam und Fachbereich jederzeit sichtbar und überprüfbar.

Empfehlungen

Der gewählte Ansatz, anhand konkreter Umsetzungen mehr über die Potentiale von Data Mesh im Unternehmen zu lernen und gleichzeitig die bestehenden Herausforderungen frühzeitig zu identifizieren, ist voraussichtlich gut auf Unternehmen verschiedener Größen übertragbar.

Das konkrete Nachweisen der Ownership-Herausforderungen im Bereich Dateninfrastruktur und Deployments war als Argumentationsgrundlage für die Initiierung der entsprechenden Änderungen sehr hilfreich. Mittelfristig werden die Angebote der Konzern-Entwicklungsplattformen entsprechend angepasst werden.

Die Data Mesh-Transformation durch konkrete Umsetzungen in begrenzten Use Cases innerhalb der bestehenden Datenlandschaft zu initiieren, hat sich als nachhaltiger herausgestellt, als mit großem Aufwand ein umfangreiches Projekt auf der grünen Wiese zu starten. Denn die so gewonnenen Erkenntnisse lassen sich leichter auf den Rest des Unternehmens übertragen und die notwendigen Veränderungen erscheinen auch für andere Unternehmensbereiche als erreichbar, da es jetzt vergleichbare Vorbilder gibt.

6.4 Einsatzbeispiel Eucon Group

Zwischen Silo und Plattform – wie Eucon mithilfe des Data Mesh-Konzepts Synergien zwischen verschiedenen Geschäftsbereichen schafft

Steckbrief

Eucon ist ein digitaler Pionier der Daten- und Prozessintelligenz. Wir kombinieren Marktdaten und Expertenwissen mit modernen Technologien wie Künstlicher Intelligenz und Robotic Process Automation zu intelligenten Automatisierungslösungen. Als Digitalisierungspartner ist Eucon auf die Branchen Automotive, Versicherungen und Real Estate spezialisiert.

Ausgangslage

Der Ist-Zustand der Eucon Group lässt sich wie folgt beschreiben: Technisch werden Daten sowohl on-premise als auch in der Cloud gespeichert. Technologie wird in den Geschäftsbereichen individuell zur Erfüllung des jeweiligen Zwecks implementiert und jeweils durch eine unterschiedliche Organisationsstruktur betreut.

Herausforderungen

Getrieben durch das Geschäftsmodell und die DNA der Eucon-Gruppe rechnen wir in den kommenden Jahren mit einem deutlich höheren Datenaufkommen. Um diesem Datenaufkommen begegnen zu können, ist es für uns elementar, die Organisation präventiv auf zukünftige Herausforderungen einzustellen. Zentrale Anforderungen können unter folgenden Punkten zusammengefasst werden:

Komplikationen bei der Interoperabilität von Daten

Sowohl die technische als auch die logische Interoperabilität stellt sich bei zunehmenden Datenmengen schwierig dar. Dies äußert sich nicht nur in der technischen Dokumentation, sondern auch in dem Verständnis von Daten für deren Nutzung. Hier soll das »Data as a Product« und das »Self Service« Prinzip des Data Meshs Abhilfe schaffen.

Erhalt der Flexibilität bei Architektur- und Technologieauswahl

Um Kundenbedürfnisse optimal befriedigen zu können, ist es von großer Bedeutung, jedem Geschäftsbereich Flexibilität zu gewährleisten. Rein zentralisierte Konzepte zum Datenmanagement würden nicht die betriebliche Realität abbilden können. Hierzu soll das Prinzip der »federated governance« zur Lösung beitragen.

Lösung

Die Vision, die wir mit der Einführung des Data Mesh-Ansatzes verfolgen, bezieht sich in erster Linie auf die Interoperabilität und dadurch mögliche Demokratisierung von Daten über eine Data Analytics Plattform. Wir haben den Ansatz des Data Mesh ausgewählt, um nicht dem Irrglauben zu verfallen, mit einer rein technischen Lösung

Kontakt

Niklas Lümmen

Data & Technology Manager

Eucon Group

niklas.luemmen@eucon.com

diese Interoperabilität herstellen und aufrecht erhalten zu können. Wir sind uns bewusst, dass es insbesondere in dem Bereich des Datenmanagements keine Silver Bullets gibt. Vielmehr haben wir festgestellt, dass es für eine generelle und langfristige Verfügbarkeit von hochqualitativen Daten sowohl technischer als auch organisatorischer Lösungen bedarf. Hierin sehen wir den größten Vorteil des Data Meshs. Das Data Mesh-Konzept bietet uns einen theoretischen Rahmen anhand dessen wir unsere technisch-organisatorischen Maßnahmen zur Etablierung dieser Data Analytics Plattform gestalten und damit die Chance einer erfolgreichen Implementierung erhöhen. Zusätzlich dazu haben wir ein iteratives Vorgehensmodell gewählt, in dem wir für eine Iterations-BU die Governance Komponente gemeinsam erarbeiten und verproben können. So schaffen wir ein frühes operatives Buy In, von dem wir uns ein steigendes Commitment für das Thema erhoffen.

Aktuell befinden wir uns bei der Implementierung des Data Mesh in einer sehr frühen Projektphase. Nichtsdestotrotz begegnen wir schon jetzt den einen oder anderen Herausforderungen. So beschäftigt uns aktuell insbesondere die Frage des »wie«. Wie setzen wir welche Tools ein, um die Implementierung des Data Mesh bestmöglich zu unterstützen? Wie viel Freiheit ist im Umgang unter Compliance-Gesichtspunkten möglich? Wie gestalten wir eine Architektur, die uns die erhofften Vorteile ermöglicht? Diese Fragen sind dabei nur ein kleiner Teil derer, an dessen Lösung wir gerade arbeiten.

Mehrwert

Durch die Einführung des Data Mesh erhoffen wir uns eine Verbesserung sämtlicher Dimensionen des DAMA Wheels. Einen besonderen Fokus legen wir dabei auf die Punkte »Data Integration & Interoperability«, »Data Security«, »Metadata«, »Data Quality« und »Data Architecture«. Die solide Basis, die wir in verschiedenen Teilen der Gruppe vorfinden, möchten wir somit ausbauen und eine Kultur etablieren, die das voneinander Lernen in den Vordergrund stellt. Diese Maßnahmen sollen langfristig zu mehr Transparenz in unseren Prozessen, zu operativer Exzellenz und übergreifend zu großem Vertrauen in unsere Daten führen. Als Unternehmen, dessen Ursprung von datengetriebenen Geschäftsmodellen geprägt ist, ist dieser Schritt für uns von enormer Bedeutung, um auch in Zukunft unseren Kunden die gewohnte Qualität und Unterstützung bieten zu können.

Empfehlungen

Eine zentrale Empfehlung, die wir trotz des recht frühen Projektstatus aussprechen können, ist die Etablierung und Beachtung eines geeigneten Change Managements. So sind wir aktuell mit einer sehr hohen Geschwindigkeit unterwegs, die uns bei einem fehlenden Change Management nicht möglich wäre. Wir legen innerhalb des Projektteams und der Iterations BU großen Wert auf die Kommunikation des übergeordneten Ziels und der gemeinsamen Erarbeitung eines Data Mesh-Konzepts, das sich an unseren Bedürfnissen orientiert. Zusätzlich achten wir darauf, keine Etablierung blinder Theorie vorzunehmen, sondern mit Bedacht und immer an der jeweiligen Realität zu agieren. So schaffen wir es, uns unserem Ziel schrittweise zu nähern und den langwierigen Prozess eines soziotechnologischen Wandels anzustoßen und mit Nachdruck voranzutreiben.



6.5 Einsatzbeispiel PwC

The PwC Data Ecosystem – A Data Mesh Approach/System:
Connect and integrate with platform & service providers

Steckbrief

PricewaterhouseCoopers GmbH WPG

- Wirtschaftsprüfungsgesellschaft, Chief Data Office
- Das Chief Data Office arbeitet an einer transparenten und ganzheitlichen Koordination von Daten, die effizientere Prozesse ermöglicht und somit PwC und ihre Kunden vorantreibt.

Ausgangslage

Da die Menge an produzierten Daten weltweit jährlich zunimmt, wird es immer schwieriger die richtigen Insights – effizient und zeitnah – aus diesen Daten zu ziehen. Bei PwC nehmen wir uns dieser Herausforderung an, indem wir skalierbare Datenprodukte und Services entwickeln, die unseren Kunden helfen Daten bestmöglich in Added-Value zu transformieren. Unser Ansatz für dieses neue Themenfeld ist es Daten selbst als Produkte zu verstehen. Diese Anstrengungen stellen eine Ergänzung unseres bisherigen Geschäft dar und helfen uns schneller, skalierbarer und effizienter auf die neuen Anforderungen des Marktes einzugehen.

Herausforderungen

Vor allem aus Perspektive unserer Software Product Owner gab es folgende Herausforderungen dabei, unsere Orientierung auf Datenprodukte noch effizienter und skalierbarer zu machen:

- Der wiederkehrende Aufwand zur Datenintegration war hoch, da vorhandene zentrale Datenverarbeitungssysteme/-Plattformen bisher einen anderen Fokus hatten. Die klare und eindeutige Trennung von Mandantendaten und verwendbaren Daten muss und wird weiterhin strikt eingehalten.
- Überlappende Tätigkeiten der Produktteams erschweren die Entwicklung von skalierenden Produkten.
- Die Zeit bis zur Markteinführung sollte schneller erfolgen und die Kunden enger in die Entwicklung eingebunden werden.
- Änderung des Schwerpunkts: Zu großer Fokus auf technische Umsetzungsdetails anstatt auf die Kundenbedürfnisse.
- Hohe Datenvarianz, verschiedene Strukturen und Formate (heterogene Datenstruktur) führten zu ineffizienter Umsetzung und erhöhten Wartungsbedarfen.
- Zentrale Datenpflege hat sich als limitierender Faktor herausgestellt
- Auffindbarkeit von Daten für andere potenzielle Konsumenten und damit Wiederverwendbarkeit war eingeschränkt.

Kontakt

Goran Loncar

Senior Manager at Chief Data Office –
Digital Products & Assets
PwC
goran.l.loncar@pwc.com

Ronny Garz

Manager at Chief Data Office
PwC
ronny.garz@pwc.com

Lösung

Das Data Ecosystem von PwC ist die zentrale Anlaufstelle für die Suche und das Auffinden von Daten(produkten). Es bietet die Möglichkeit zur Integration verschiedenster Datenquellen, die zu Daten-Produkten kombiniert und für unsere Software-Produkte und Dienstleistungen bereitgestellt werden. Für die Verwaltung der Daten ist es nicht notwendig die Daten physisch im Data Ecosystem zu speichern, da es auch in der Lage ist Daten und Metadaten aus verteilten Systemen zu nutzen. Dies ermöglicht es, Governance-Fähigkeiten an einer zentralen Stelle unabhängig der Datenablage zu etablieren. Die Daten selbst werden über (semi-)automatisch gepflegte Kataloge sichtbar und auffindbar gemacht. Durch die Nutzung zentraler und standardisierter Schnittstellen und Blaupausen wird das Onboarding neuer Produkte und Dienste stark vereinfacht, die Skalierbarkeit erhöht sich. Neben dem eigentlichen Datenprodukt bietet das Data Ecosystem verschiedene Dienste – wie Data Quality Monitoring oder Lineage –, die es Data Ownern ermöglichen ihren Kunden und Nutzern die Daten nachvollziehbar und in höchster Qualität anzubieten.

Mehrwert

-Auszug-

Das PwC Data Ecosystem bietet einen technologischen Rahmen für

1. eine vereinfachte Entwicklung von Datenprodukten,
2. eine verkürzte Markteinführungszeit durch angepasste, domain/branchenspezifische Datenprodukte
3. die Bereitstellung von Informationen zu Datenprodukten in einem Datenkatalog für die PwC-Produktentwicklungsteams
4. eine zentrale Governance und Übersicht über die vorhandenen Datenprodukte zur Unterstützung der End-to-End-Compliance
5. die Befähigung der Daten(produkt)entwicklungsteams, sowohl mit strukturierten als auch mit unstrukturierten Daten, Batch und Streaming zu arbeiten
6. eine einfache Anknüpfung an andere Daten- und Cloud-Plattformen durch eine virtuelle Speicherebene
7. Das Erkunden von Daten und Trainieren, Bewerten und Bereitstellen von ML-Modellen für Analysefunktionen (Analytics & Data Science at scale) an zentraler Stelle ermöglicht es uns, weitere Use Cases durch neue Einblicke in die Daten zu finden, zu realisieren und somit neue Produkte und Dienste anbieten zu können

Empfehlungen

Lessons Learned und Empfehlungen:

- Data Mesh setzt vor allem ein Umdenken voraus und erfordert umfangreiche Organisationsänderungen. Diese stellen tief etablierte Muster in Frage, bspw. »Wem gehören welche Daten?« »Wer ist verantwortlich für ...?«
- Da es sich um ein strategisches Projekt handelt, muss der Wille und die Unterstützung von Organisation und Unternehmensführung bei der Umsetzung gegeben sein, um Organisationsänderungen auch erfolgreich durchführen zu können.
- Die Domain-Teams benötigen ausreichendes Wissen und müssen aus-/weitergebildet werden, um das Verständnis für Konzepte rund um Qualität, Standards und

notwendige Verantwortung zu erlangen. Auch hier braucht es ein Umdenken weg von Software als Produkt hin zu Daten als Produkt.

- Wir empfehlen mit einzelnen und möglichst konkreten Anwendungsfällen zu beginnen, bei denen sich ein Mehrwert und Ziel bereits absehen lässt. Dies könnte bspw. die Etablierung eines Datenprodukts sein, die mehrere separate Implementierungen vermeidet.
- Zur effizienten Erstellung von Blaupausen und Templates ist es wichtig, existierende Datenprodukte miteinander zu vergleichen und Gemeinsamkeiten gegenüber neuen Datenprodukten zu validieren.

7 Einführung und Umsetzung

Je nach Aufgabenstellung ergeben sich zur Einführung und zur erfolgreichen Umsetzung von Data Mesh in Unternehmen und Organisationen mehrere Herangehensweisen. Eine mögliche Herangehensweise unter vielen ist das evolutionäre und inkrementelle Vorgehen.

Dieses Vorgehen schafft durch fokussierte und schrittweise Veränderungen der Organisation (z.B. nur in einer ausgewählten Abteilung oder anhand isolierter Anwendungsfälle) erste schnelle und sichtbare Erfolge. Gleichzeitig werden die Auswirkungen eines möglichen Scheiterns der Umsetzung reduziert. Dabei werden vier Stufen unterschieden:

Definition der Data Mesh Strategie & Chancenexploration

- Formulierung einer Data Mesh Vision & Mission passend zu den Unternehmenszielen
- Identifikation & Bewertung möglicher Data Mesh Handlungsfelder (z.B. Umsetzung innerhalb einer Organisationseinheit wie Marketing) und Anwendungsfällen (z.B. einem Marketingreport für Kundenzufriedenheit)

Erbringung eines Data Mesh Wertnachweises (Proof of Value) & Wirksamkeitsnachweises (Proof of Concept)

- Erstellung einer Business Cases Kalkulation mit indikativen Kosten und erwarteten Mehrwerten der Data Mesh Initiative, um notwendige Sponsoren zu überzeugen (Proof of Value)
- Validierung der angenommenen Mehrwerte und Beweis der Data Mesh Wirksamkeit durch prototypische und fokussierte Implementierung (Proof of Concept) eines Anwendungsfalles in einer kontrollierten Umgebung (z.B. Datenversorgung für einen Marketingreport für Kundenzufriedenheit)

Implementierung & Pilotierung eines minimal funktionstüchtigen Data Mesh Produktes (Minimal Viable Product)

- Weiterentwicklung des Proof of Concepts zu einer ersten minimal funktionstüchtigen Version des Produktes und Pilotierung mit ausgewählten Nutzern (z.B. Kundenzufriedenheitsreport für ein ausgewähltes Kundensegment)
- Sammlung und Auswertung von Nutzerfeedback, um schnell für die weitere Entwicklung des finalen Data Mesh Ansatzes zu lernen und bei Fehlentwicklungen früh gegenzusteuern

Skalierung und Erweiterung zur finalen Data Mesh Organisation

- Ausweitung des MVP zu einem finalen und umfassenden Data Mesh und Definition des Organisationsmodells (z.B. Umsetzung für alle Reports in der Marketingabteilung)
- Inkrementelle Erweiterung auf andere Handlungsfelder (z.B. weitere Abteilungen neben Marketing) und auf weitere Anwendungsfälle unter Wiederverwendung eines inkrementellen vierstufigen Ansatzes.

Bitkom vertritt mehr als 2.000 Mitgliedsunternehmen aus der digitalen Wirtschaft. Sie erzielen allein mit IT- und Telekommunikationsleistungen jährlich Umsätze von 190 Milliarden Euro, darunter Exporte in Höhe von 50 Milliarden Euro. Die Bitkom-Mitglieder beschäftigen in Deutschland mehr als 2 Millionen Mitarbeiterinnen und Mitarbeiter. 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 oder in anderer Weise Teil der digitalen Wirtschaft. 80 Prozent der Unternehmen haben ihren Hauptsitz in Deutschland, jeweils 8 Prozent kommen aus Europa und den USA, 4 Prozent aus anderen Regionen. 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 weltweit führenden Digitalstandort zu machen.

Bitkom e.V.

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

[bitkom.org](https://www.bitkom.org)

bitkom