

# DATA MESH

## PRINZIPIEN, NUTZEN UND IMPLEMENTIERUNG

von Alexander Hofstetter

### IN KÜRZE

- Bei der Implementierung eines Data Mesh werden Daten als Produkt gesehen, ihr Besitz geht in Datendomänen über, es gibt eine zentrale Self-Service Infrastruktur und die Governance ist mehrheitlich Sache der Datendomänen.
- Diese Struktur verbessert sowohl die Qualität als auch die Auffindbarkeit der Daten.
- AWS kann Unternehmen dabei helfen, eine zentrale Infrastruktur zu etablieren, über die Datendomänen ihre Daten zur Verfügung stellen und die Zugriffe auf diese erteilen können.



Episode schauen  
[trivadis.com/sparx](https://trivadis.com/sparx)



Aktuell kommt man kaum am Thema Data Mesh vorbei. Es ist ein Paradigmenwechsel in der Art und Weise, wie wir über den Aufbau von Datenplattformen denken, als datengetriebenes Unternehmen Daten organisieren und gleichzeitig potenzielle Probleme mit bereits etablierten Technologien – wie sie z. B. für einen zentralen Data Lake eingesetzt werden – lösen können.

Begründet wurde das Thema von Zhamak Dehghani<sup>1</sup>, Technology Principal bei ThoughtWorks, die auch die vier Design-Prinzipien eines Data Mesh beschreibt.

<sup>1</sup> <https://martinfowler.com/articles/data-mesh-principles.html>  
<https://martinfowler.com/articles/data-monolith-to-mesh.html>

# WELCHE DESIGN-PRINZIPIEN GIBT ES?

Es gibt vier verschiedene Säulen für ein Data Mesh:

- Domain-Driven Data Ownership Architecture
- Data as a Product
- Self-Service Infrastructure as a Platform
- Federated Computational Governance

## DOMAIN-DRIVEN DATA OWNERSHIP ARCHITECTURE

Bei der Domain-Driven Data Ownership Architecture geht es um den Wechsel von zentralen Datenplattformen mit zentraler Datenverantwortung zu einem Ansatz, in dem der Besitz der Daten in sogenannte Datendomänen übergeht – vergleichbar zum Applikationsdesign.

Eine zentrale Datenplattform stellt einen großen Monolithen dar, der alle Bereiche der Applikation abdeckt. Im Gegenzug hat man viele kleine Microservices implementiert, die jeweils nur eine bestimmte Funktion abdecken. Beim Data Mesh ist die Datendomäne der kleine Microservice und die zentrale Datenplattform (bspw. ein Data Lake oder ein Enterprise Data Warehouse) der Monolith.

Die Datendomänen bestehen aus Teammitgliedern, welche auch die fachliche Expertise über die Daten besitzen. Dadurch können sie besser bestimmen, welche Daten gesammelt werden, wie sie transformiert oder aggregiert werden und wie man sie am besten weiterverarbeitet. Im Data Mesh ist es auch möglich, dass der/die Datenproduzent\*in gleichzeitig Konsument\*in der eigenen Daten ist. Man verteilt die Daten also an der Quelle und geht nicht den Umweg über eine zentrale Einheit. Der Schlüssel liegt in der Betrachtung der Daten als Produkt.

**« Der Schlüssel liegt in der Betrachtung der Daten als Produkt.»**

Wenn Konsument\*innen Daten benötigen, gehen diese mit ihren Anforderungen nicht wie bisher an zentrale BI/Data Analytics oder Data Platform Teams, sondern direkt auf die Data Owner zu. Der Vorteil davon ist, dass man die zentralen Plattform-Teams mit dem Aufbau neuer Transformations- oder Bereitstellungsprozesse entlastet. Auch kann sich der/die Konsument\*in direkt mit dem/ der Produzent\*in austauschen, also der Instanz, welche die Daten am besten kennt. Um zu verhindern, dass dadurch neue Datensilos entstehen, bekommen die Datendomänen auch die Verantwortung, ihre Datenprodukte zu vertreiben sowie ausführlich und einheitlich zu dokumentieren.

## DATA AS A PRODUCT

Innerhalb der Datendomänen wird die neue Rolle Data Product Owner geschaffen. Dieser ist verantwortlich dafür, die Daten als Produkt (Data as a Product) auszuliefern. Dabei ist es wichtig, die Daten genau zu beschreiben, damit Datenanalysten wie auch andere Datendomänen sie finden. Zusätzlich steht der Data Product Owner in der Verantwortung für die Qualität und das Wachstum der Daten. Der Data Product Owner weiß, wie viele Konsument\*innen seine Daten nutzen und wie zufrieden diese sind.

## SELF-SERVICE INFRASTRUCTURE AS A PLATFORM

Die dezentralen Data Owner benötigen eine zentrale Infrastruktur als Plattform für ihre Daten. Hier ist es besonders wichtig, dass die Infrastruktur domänenagnostisch bleibt und keine Verantwortung in eine zentrale Komponente übergeht. Das bedeutet, dass die Teams die Datenplattform nutzen und über diese selbst steuern können, welche Daten sie zur Verfügung stellen und wer auf diese zugreifen kann.

Bei Self-Service Infrastructure as a Platform wird entsprechend nur ein Self-Service zur Verfügung gestellt, um diese zentrale Infrastruktur zu nutzen. Das hat den Vorteil, dass die Datendomänen die gleichen Standards bei der Infrastruktur nutzen, ohne vorher mit einer zentralen Einheit zusammenzuarbeiten. Diese wird dadurch entlastet.

## FEDERATED COMPUTATIONAL GOVERNANCE

Das letzte Prinzip sorgt dafür, dass die unterschiedlichen Datenprodukte auch sinnvoll genutzt werden können. Data Mesh verfolgt grundsätzlich einen verteilten Ansatz – Federated Computational Governance trägt dem Rechnung, indem die Governance nicht nur ausschliesslich zentral aufgesetzt und gelebt wird, sondern viel Verantwortung an die Datendomänen delegiert wird.

Entscheidungen, die das Innenleben einer Domäne betreffen, sollen auch von der Domäne getroffen werden. Nichtsdestotrotz braucht es auch eine zentrale Governance, welche die Themen behandelt, die von übergreifendem Interesse sind, wie z. B. Standardisierung der Schnittstellen, Nutzung eines Datenkatalogs, sowie die Verwaltung von Zugriffen unter Einhaltung der eigenen Compliance- und Berücksichtigung der Security-Regeln.

**« In Sachen Governance wird viel Verantwortung an die Datendomänen delegiert: Entscheidungen, die das Innenleben einer Domäne betreffen, sollen auch von der Domäne getroffen werden.»**

## IST DATA MESH EINE LÖSUNG FÜR MICH?

Neben den vielen Mehrwerten – z. B. Datenzugehörigkeit sowie Verantwortlichkeit und die Möglichkeit, Business-Anforderungen an die Lösung zu knüpfen und flexibel an Organisationsstrukturen zu skalieren – gibt es auch Herausforderungen bei der Implementierung eines Data Meshs.

So benötigt man für ein Data Mesh verteiltes Wissen und Skills in den Datendomänen. Zudem besteht die Gefahr von voneinander abweichenden Technologie-Stacks. Produktdenken und die Erstellung von Daten als Produkt ist des Weiteren ein komplexer Vorgang, der von allen beteiligten Mitarbeitenden getragen werden muss.

Die Tragweite der Etablierung eines Data Meshs kann mit der Umstellung auf DevOps verglichen werden – es geht also nicht nur um die Implementierung einer neuen Datenplattform.

Deshalb ist es essenziell zu prüfen, welche Hürden überwunden werden müssen, um ein Data Mesh zu nutzen. Fragen, die sich dabei stellen, sind:

- Wie kann ich bestehende Datenplattformen auseinanderbrechen und in ein Data Mesh integrieren?
- Kann ich das Mindset, Daten als Produkt zu sehen, im Unternehmen etablieren?
- Habe ich Mitarbeitende, die als Data Product Owner agieren und die Daten dokumentieren, bewerten und vermarkten können? Und auch dafür verantwortlich gemacht werden können?

Letztlich hängt der Implementierungsprozess auch davon ab, ob man Datenplattformen bereits in Betrieb hat oder erst startet.

## WIE KANN AWS MIR DABEI HELFEN?

Während man die Verschiebung zu Data-Driven Domains und die Behandlung von Daten als Produkt alleine bewerkstelligen muss, hat AWS Werkzeuge, die es unter Einhaltung von Governance-, Compliance- und Security-Aspekten ermöglichen, eine Infrastruktur als Plattform bereitzustellen. Zu den wichtigsten gehört ein Managed Data Lake Service namens AWS Lake Formation.

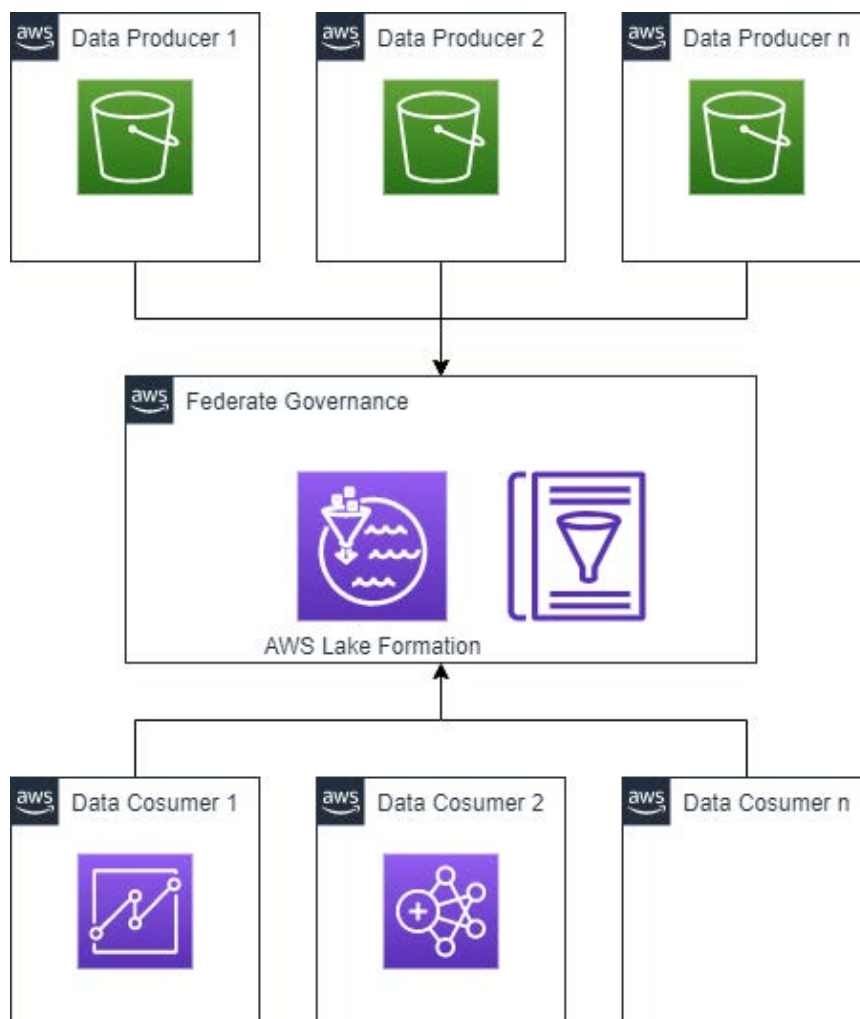
Mit AWS Lake Formation können Datendomänen ihre Daten einfach in einen Data Lake migrieren, transformieren, bereinigen und katalogisieren, um diese dann für Datenkonsument\*innen zur Verfügung zu stellen.

AWS Lake Formation bietet zudem eine einheitliche Autorisierungsschicht, um den Zugriff auf die Daten zu gewährleisten. Ein weiterer wichtiger Aspekt ist, dass die Daten nicht erst zu den Konsument\*innen kopiert werden müssen, sondern direkt abgerufen werden können.

Somit hat man eine zentrale Plattform für die Datendomänen, die ihre eigenen Daten beliebigen Konsument\*innen zur Verfügung stellen und deren Zugriff steuern können. Der Zugriff erfolgt hierbei über APIs und die Autorisierung über ein Security Token.

## WIE SIEHT EINE AWS LAKE FORMATION ARCHITEKTUR AUS?

Datenproduzent\*innen registrieren ihre Daten in AWS Lake Formation. Konsument\*innen können im Anschluss Zugriff auf die Daten anfordern und bei Freigabe direkt nutzen.



# WIE IST DER ZUGRIFF GESTALTET?

Nachfolgend ist ein Workflow vom Registrieren der Daten durch den/ die Datenproduzent\*in bis zur Freigabe und Nutzung durch den/ die Konsument\*in abgebildet:



Bildquelle: <https://aws.amazon.com/blogs/big-data/design-a-data-mesh-architecture-using-aws-lake-formation-and-aws-glue/>

## FAZIT

Die Implementierung eines Data Meshs geht weit über die Technik hinaus und kann viele Vorteile bringen. Für die technische Implementierung bietet AWS Werkzeuge, die es Datendomänen erlauben, eine einfache und sichere Umgebung zu nutzen, um ihre Daten zur Verfügung zu stellen sowie die Zugriffe auf diese zu erteilen. Damit bleibt die Verantwortung in den Teams und die Datenkonsument\*innen können direkt mit den Besitzer\*innen der Daten reden. Durch die neu geschaffene Rolle des Data Product Owners wird die Qualität der Daten verbessert. Dank guter Dokumentation werden diese zudem leichter von potentiellen Datenkonsument\*innen gefunden.

## ÜBER DEN AUTOR

### Alexander Hofstetter

Alexander Hofstetter ist als Principal Consultant bei Trivadis – Part of Accenture in München im Bereich Infrastructure Managed Services tätig. Einer seiner Schwerpunkte liegt in der Oracle-Datenbank-Administration. Neben Automatisierung beschäftigt er sich seit 2017 mit der Administration von Big Data Clustern. 2018 kam das Design von AWS Cloud Architekturen, sowie der automatisierte Aufbau von Infrastrukturen in der AWS Cloud mit Infrastructure as Code dazu. Aktuell arbeitet er als Product Owner im Bereich AWS im Cloud Competence Center.

EMAIL [alexander.hofstetter@trivadis.com](mailto:alexander.hofstetter@trivadis.com)

TELEFON +49 89 9927 59 302

SPARX.TRIVADIS.COM



trivadis Part of Accenture

