

User Story – Data Spaces

Stand 2022

Ziel des Dokuments:

Die Notwendigkeit und Funktion von Data Spaces anhand eines Beispiels verständlich machen.

Inhaltsverzeichnis

<i>Einleitung</i>	2
<i>User Story pt. I</i>	6
<i>User Story pt. II</i>	14

Autor*innen:

Natascha Totzler (natascha.totzler@gmail.com) und

Günther Tschabuschnig, DIO-Präsident (guenther.tschabuschnig@gmail.com)

Das vorliegende Werk wurde unter CC by 4.0 AT veröffentlicht

Abstract

Die Menschheit generiert, misst und managt täglich Milliarden neuer Daten: Umweltdaten, Gesundheitsdaten, Daten aus sozialen Medien, u.v.m. Bei der Nutzung und dem Austausch von Daten gelangen Unternehmen jedoch schnell an ihre Grenzen. Das angeführte Beispiel zeigt, wie diese Grenzen zum Nutzen für Organisationen für neue datengetriebene Businesscases führen kann.

In einer Welt, die immer stärker datengetrieben agiert und Produkte aus Daten kreiert, ist es unabdingbar auch technologisch die geeignete Umgebung dafür zu schaffen. Die dazu angeführten Werkzeuge, wie Data Spaces oder Konnektoren sind Bausteine, um ein Datenökosystem aufzubauen, zu nutzen und dabei zu partizipieren. Dabei wird besonders hoher Wert auf die Themen Souveränität, Trust und das Vernetzen von Daten gelegt.

In unserem Beispiel zum Aufbau von Data Spaces bedienen wir uns des Beispiels eines digitalen Produkt Passes, der die Nachvollziehbarkeit einer Lieferkette darstellt und dabei das Thema CO₂ genauer betrachtet, um darzustellen, wie Produktverbesserungen oder gar neue Produkte mit datengetriebener Innovation und den Bausteinen der Datenökonomie vorangetrieben werden können.

"Without Data, you're just another person with an opinion"

- W. Edwards Deming

Einleitung

Ein regionaler BIO-Supermarkt Betreiber möchte künftig mehr Awareness auf den ökologischen CO₂-Fußabdruck von Lebensmitteln bringen. Das Ziel dahinter ist die Bewusstseinschaffung bei Konsument*innen und Lieferant*innen, um auf Dauer nachhaltiger und Ressourcen schonender zu produzieren und anzubieten, da ermittelt wurde, dass die Angabe des ökologischen CO₂-Fußabdrucks die Kaufentscheidung positiv beeinflusst.

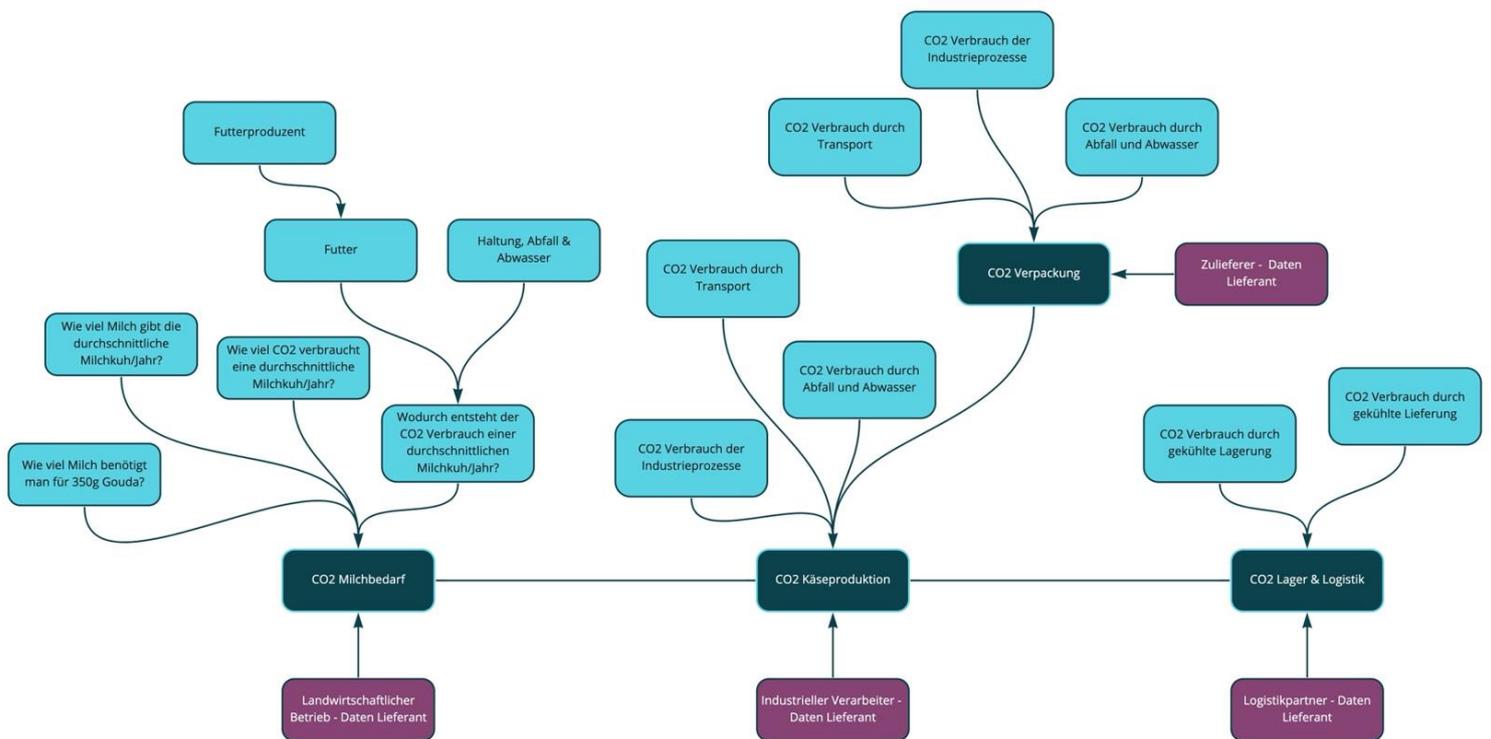
Das Produkt, das also entstehen soll, ist die datenbasierte Berechnung des CO₂-Verbrauchs jedes Lebensmittels in Form einer Plakette auf jedem Produktetikett aller Eigenmarken-Produkte.

Der/die BIO-Supermarkt Betreiber*in hat eine umfangreiche Produktpalette. Diese Produktpalette umfasst hauptsächlich Grundnahrungsmittel und Milchprodukte. Da dies ein sehr umfangreiches Vorhaben ist, beschließt das Unternehmen einen Proof of Concept anhand einer Lieferkette für ein Produkt vorzunehmen. Der allseits beliebte Gouda aus biologischer Landwirtschaft wird für diesen Proof of Concept ausgewählt.

Alex, die Projektleiterin des Vorhabens, identifiziert alle Stakeholder und ermittelt was es bedarf, um für dieses eine Produkt den CO₂-Fußabdruck zu berechnen. Eine erste Recherche ergibt folgende Eckpunkte:

- CO₂-Verbrauch für den Milchbedarf einer 350g Packung Gouda
- CO₂-Verbrauch in der Käseproduktion
- CO₂-Verbrauch für Verpackung
- CO₂-Verbrauch für Lager und Logistik bis in den Supermarkt

Da dies aber nicht nur einzelne Zahlen sind, sondern sich auch diese auf unterschiedliche Berechnungen stützen, visualisiert sie, welche Daten von wem benötigt werden:



Schnell wird ersichtlich, dass allein diese Berechnung für nur ein einzelnes Produkt sehr viele unterschiedliche Datenquellen und Daten-Lieferant*innen benötigt. Und selbst wenn sie alle Daten einmalig eingesammelt hat, wird sie dennoch Hilfe von außerhalb des Unternehmens benötigen, um die Berechnung durchzuführen.

Sie macht sich also an die Arbeit, tritt mit ihren Lieferant*innen in Kontakt und holt alle mit an Bord. Jede*r ist grundsätzlich interessiert bei dem Versuch mitzumachen, um zu sehen, ob dieses Vorhaben überhaupt realisierbar ist, auch wenn der/die eine oder andere Partner*in skeptisch ist, ob die Daten auch genauso verwendet werden, wie besprochen.

Der BIO-Landwirtschaftliche Betrieb und die BIO-Käseproduktion finden die Transparenz und die Bewusstseinsbildung gut und zusätzlich erhoffen sie sich auf Dauer eine gesteigerte Abnahme durch bessere Verkaufszahlen. Der datenaffine Logistikpartner hofft das natürlich auch, sieht aber zusätzlich das Potential datenbasierend Logistikketten zu optimieren und CO₂ einzusparen. Immerhin muss das ja nicht dauerhaft eine Datenlieferungs-Einbahnstraße sein.

Während sich die Datenlieferant*innen daran machen alle notwendigen Zahlen zu ermitteln, findet Alex den passenden Datenverarbeitungspartner (Data Intermediary¹). Ein Start-Up, das sich darauf spezialisiert hat, Daten aus unterschiedlichen Datenquellen sowohl inhaltlich und qualitativ zur Weiterverarbeitung aufzubereiten als auch die notwendigen Berechnungen durchzuführen.

Da es sich hier ganz klar um einen Proof of Concept handelt, wird ein entsprechender NDA (Non-Disclosure Agreement) aufgesetzt und Alex sorgt für einen digitalen Speicherort, auf dem die Daten abgelegt werden sollen. Aufgrund der Einmaligkeit des Vorhabens sind alle damit einverstanden ihre Daten auf dem Drive des BIO-Supermarkt Betreibers abzulegen, von wo auch das Datenverarbeitungs-Unternehmen darauf zugreifen kann.

Alle NDAs sind unterzeichnet, die Daten werden im digitalen Raum des BIO-Supermarkt Betreibers abgelegt, die Daten werden verarbeitet und voilà der Gouda hat einen CO₂-Fußabdruck.

Damit Alex den gelungenen Proof of Concept auch für das Management und zu Marketingzwecken aufbereiten kann, wendet sie sich an die Grafik-Abteilung. Die entwerfen ein entsprechendes Verpackungsdesign, wie denn die Produktetiketten künftig aussehen können.

Die Präsentation kommt gut an, Alex wird beauftragt das CO₂-Fußabdrucks-Label für alle Produkte der Eigenmarke zu entwickeln.

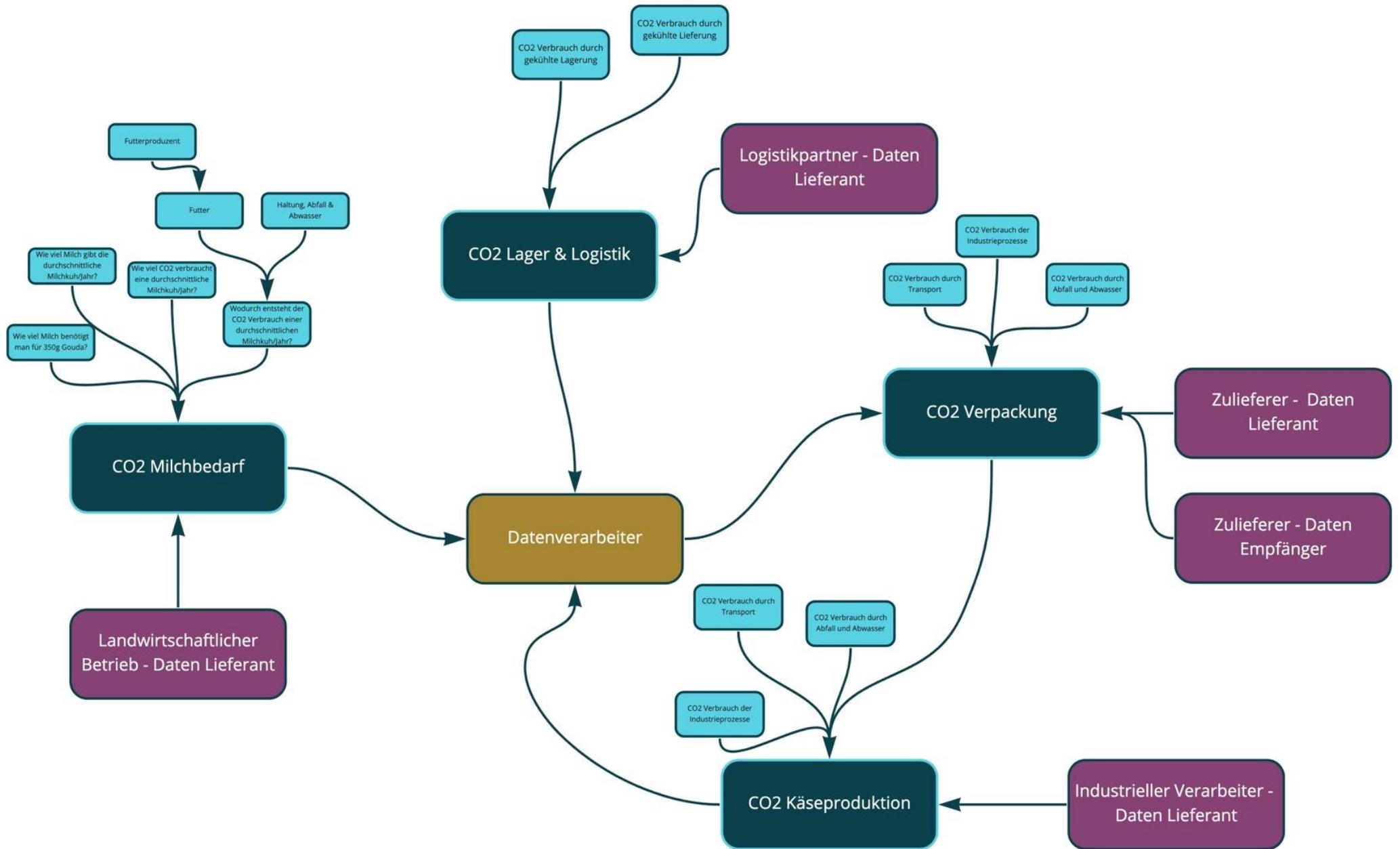
¹ Als Data Intermediary bezeichnet man eine Organisation, die Daten im Auftrag einer anderen Organisation verarbeitet. Der Data Intermediary arbeitet daran, die Daten auf bestimmte Weise zu verwalten, und bietet ein gewisses Maß an Vertrauen hinsichtlich der Verwendung der Daten.

User Story Teil I

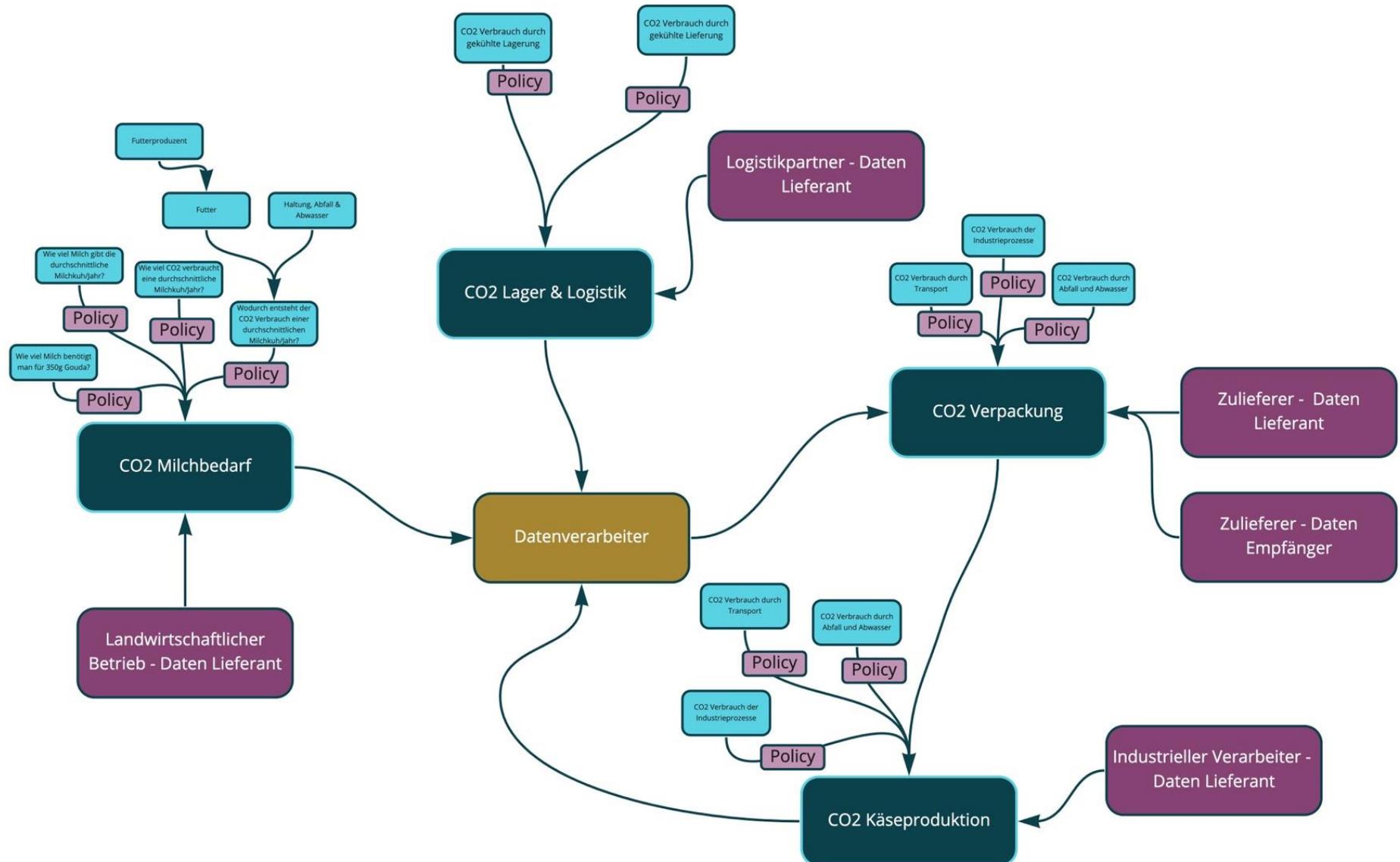
Alex steht nun vor der Herausforderung all die Lieferant*innen und Lieferketten-Teilnehmenden an Bord zu holen. Dieses Mal aber mit weitaus höheren Anforderungen als bei dem Proof of Concept, da sich durch kontinuierliche Datenströme und deutlich mehr Teilnehmende an dem künftigen datenbasierten Produkt, die Komplexität exponentiell erhöht hat.

Sie ermittelt folgende neue Anforderungen:

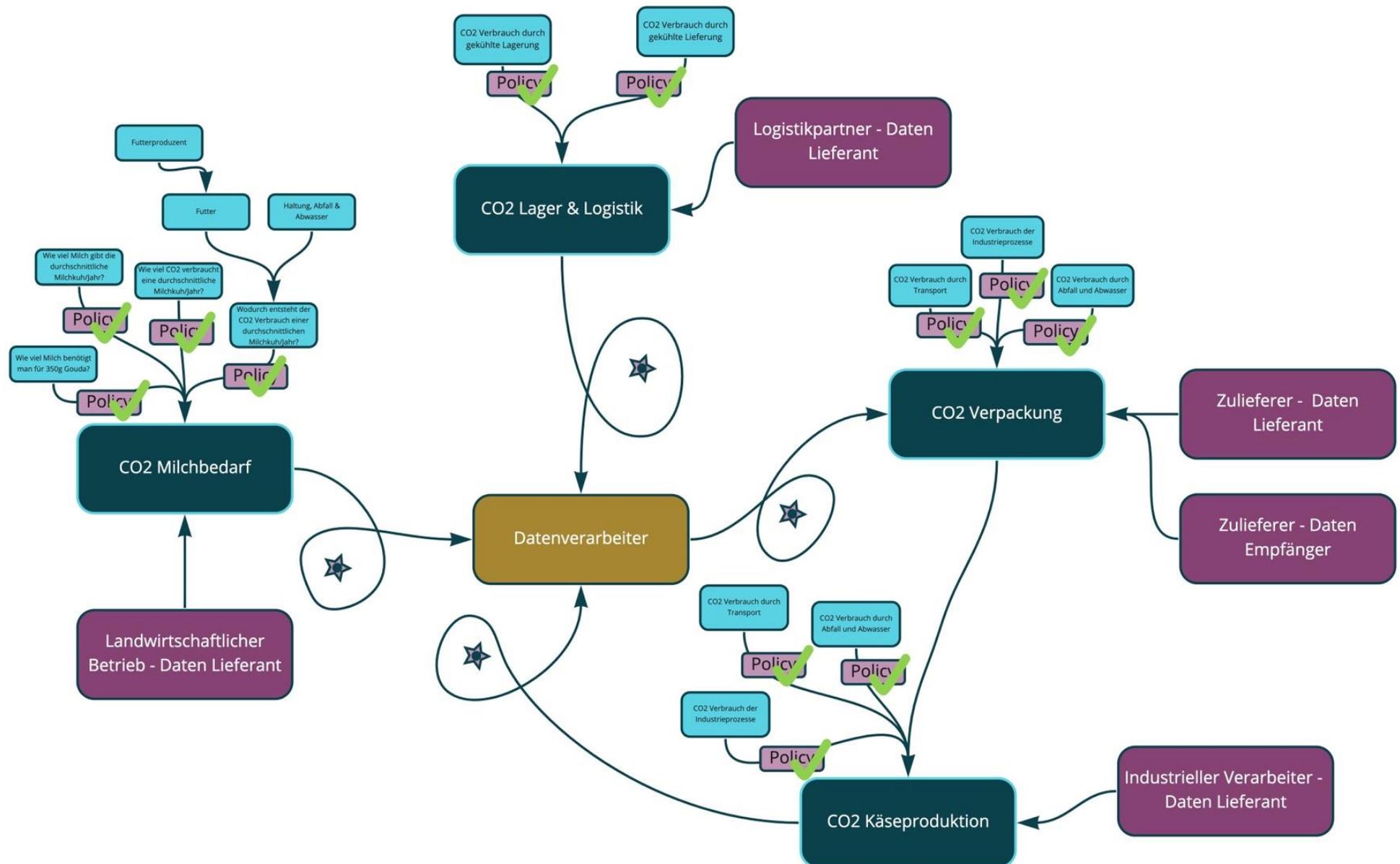
- Zunächst dürfen keine manuellen Schritte mehr für den Prozess notwendig sein, da dies nicht sinnvoll umsetzbar wäre. Das heißt unter anderem, dass die ermittelten Werte an den Verpackungsproduzenten zurückgespielt werden müssen, damit dieser die errechneten Werte auf die Etiketten drucken kann. Es müssen also Daten zum/zur Datenverarbeiter*in fließen und von dort wieder weg. Wenn man das im Falle des Goudas durchdenkt, würde dies in etwa so aussehen:



- Außerdem haben die Datenlieferant*innen auch die Anforderung gestellt, dass sie definieren können müssen, welcher Datensatz unter welchen Bedingungen (= Usage Policy) mit wem (= Access Policy) geteilt werden darf. Immerhin sind das vertrauliche Produktionsdaten, die in den falschen Händen beachtlichen Schaden anrichten können. Wenn man das im Falle des Goudas durchdenkt, würden an einigen Stellen Verwendungs- und Zugriffs-Policies hängen:

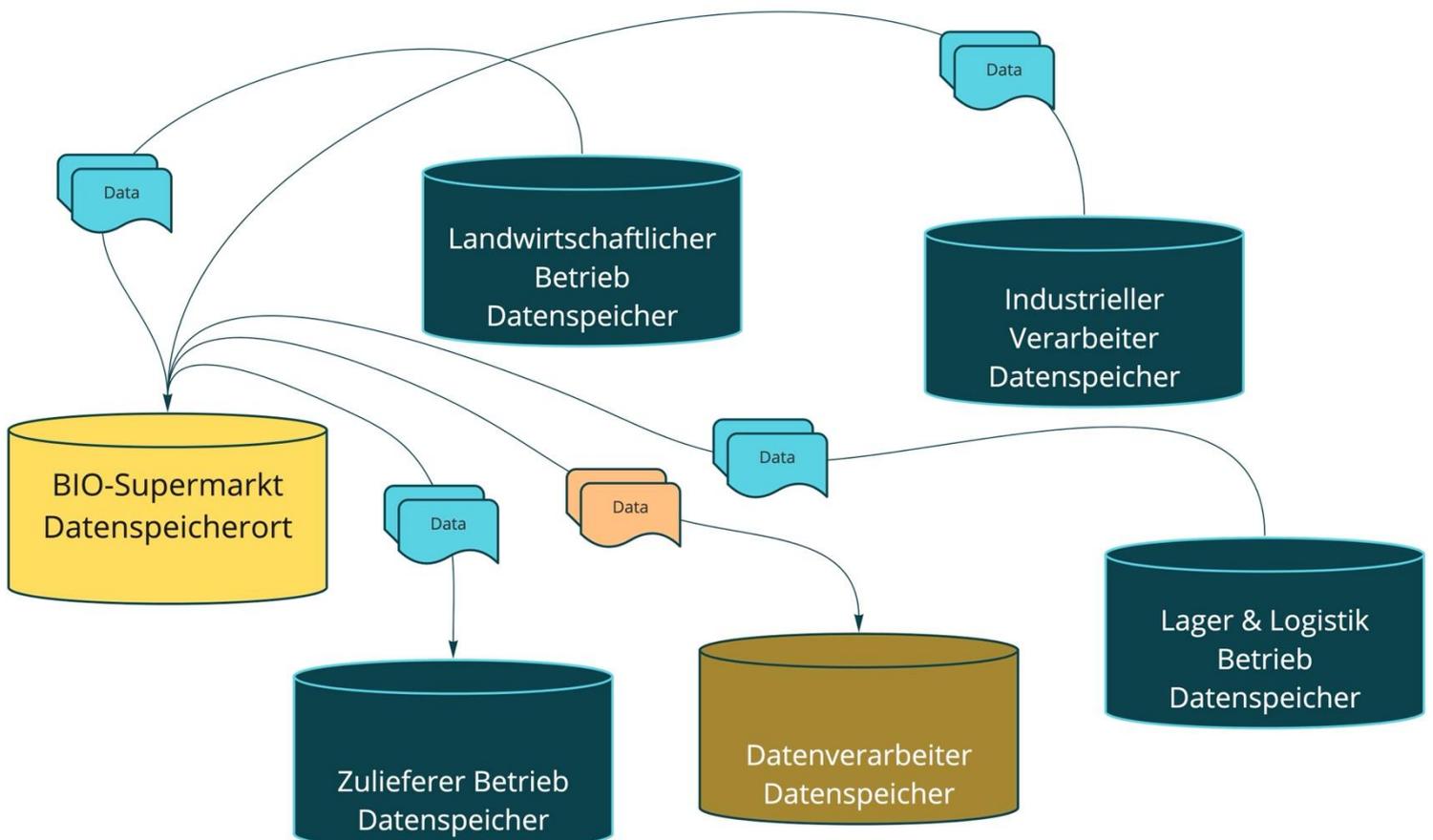


- Zusätzlich müsste der Datenaustausch automatisiert passieren und auch nur dann erfolgen, wenn die Bedingungen stimmen. Also nur dann, wenn der Datenempfänger*innen auch den Policies des/der Datenanbieter*in zugestimmt hat:

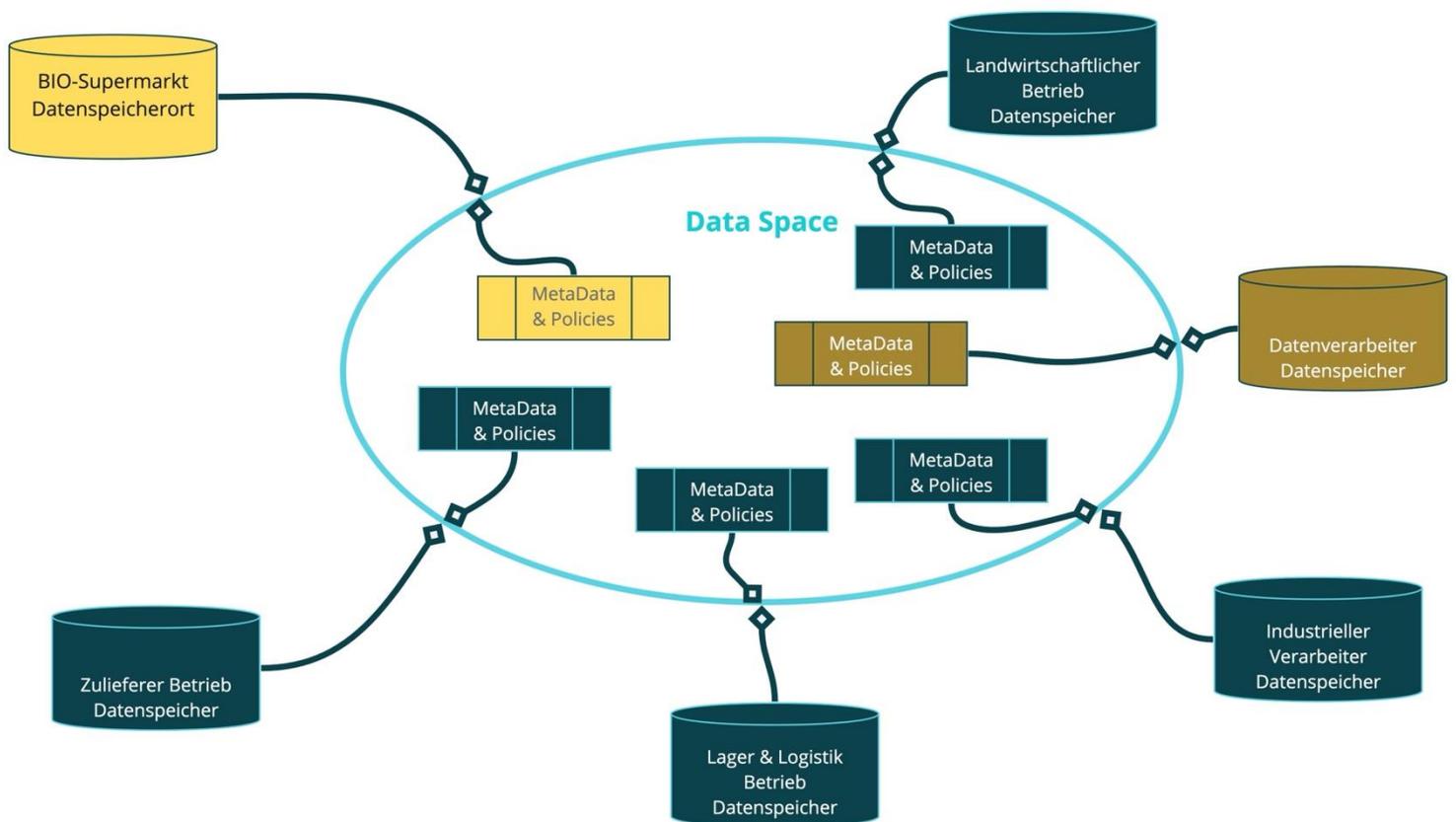


- Unglücklicherweise lässt sich das aber alles nicht voll automatisiert abbilden, da der erste Schritt, wie diese Daten gesammelt werden, das Hochladen in das Drive des BIO-Supermarkt Betreibers war. Ab diesem Zeitpunkt hat ein/eine Datenlieferant*in bereits die Kontrolle über seine/ihre Daten verloren, da sie sein/ihr eigenes Hoheitsgebiet verlassen haben (Souveränitätsproblem). Eine weitere Anforderung ist somit ein neutraler digitaler Boden, auf dem eine Verbindung hergestellt werden kann, damit man den Bedingungen zustimmen kann, ohne dass die Daten bereits ihre Quelle verlassen.

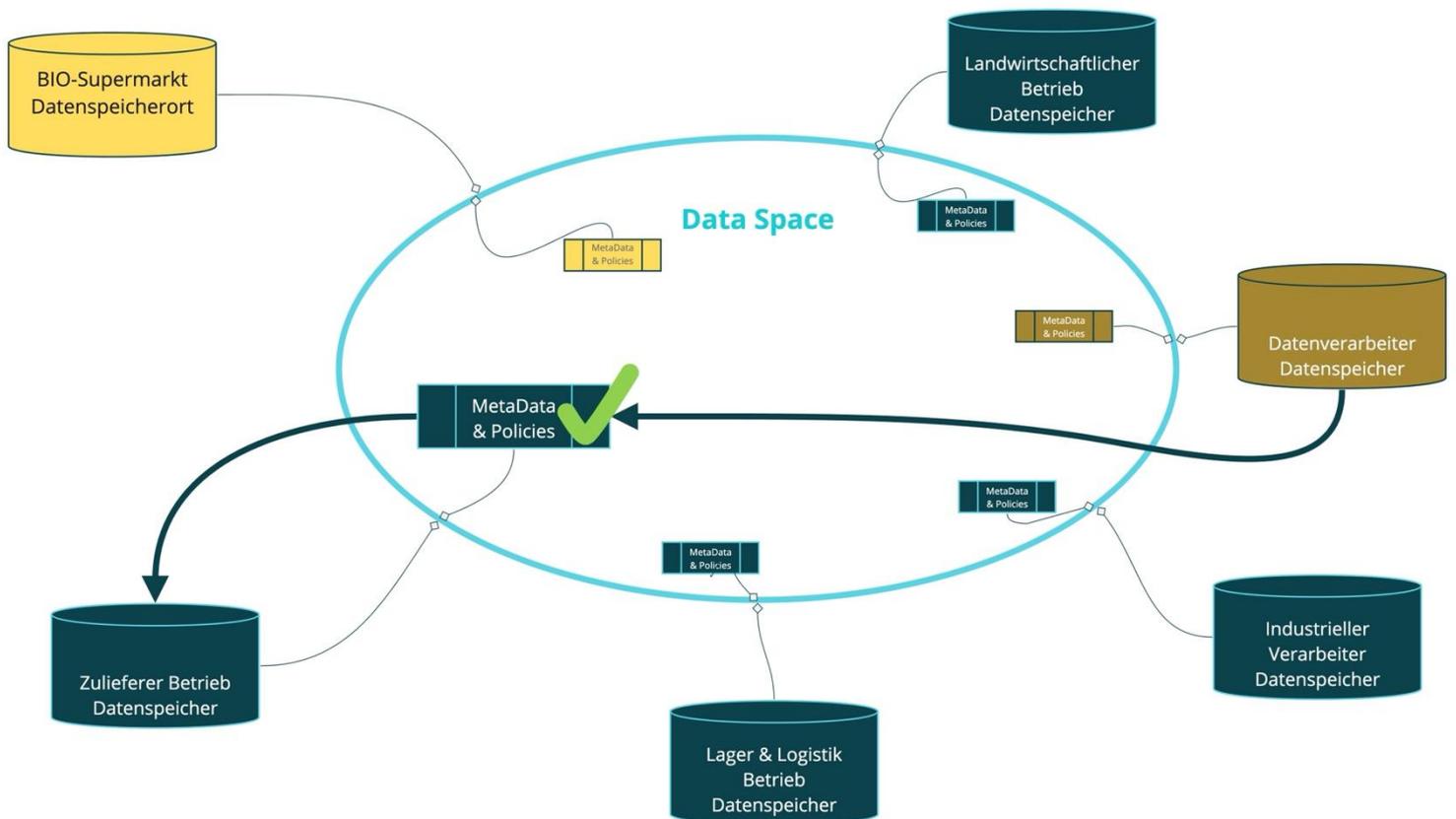
Für den Proof of Concept wurden also Daten direkt in den Speicher des BIO-Supermarkt Betreibers geliefert. In diesem Fall war das auch vollkommen in Ordnung, da alle damit einverstanden waren, es ein kleiner übersichtlicher Kreis war, die einmaligen Daten auch nicht unternehmenskritisch waren und auch ein gewisses Vertrauen zwischen den Parteien bestand, ergab sich folgendes Bild:



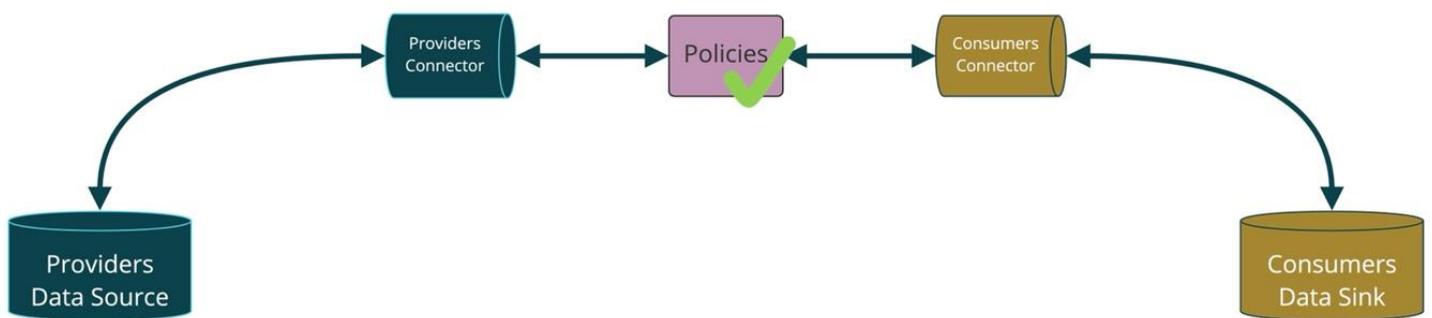
Um das datenbasierte Produkt, das CO₂-Label auf jeder Produktverpackung, aber auch im großen Stil bei allen Produkten der Eigenmarke und unter Partner*innen, die sich nicht kennen, ermöglichen zu können, benötigt es eine andere Art der Architektur und des Datentransfers. Einen digitalen Raum, in dem sich Partner*innen finden, auf Beschreibungen der Daten (= MetaDaten) zugreifen und auf Bedingungen zum Datenaustausch (= Policies) einigen können, bevor ein tatsächlicher Datenaustausch stattfindet. Dieser Raum ist ein Data Space:



Teilnehmende in einem Data Space sehen die Beschreibung von Daten, aber nicht deren Inhalt! Wie ein Buch-Cover, das einem einen Titel, Autor*in und eine Beschreibung liefert (=MetaDaten) und ein Preisschild, das die Bedingungen nennt, zu denen man auf den Inhalt zugreifen darf (= Policies). Also in unserem Beispiel könnte das ein MetaDaten-Satz sein, der als Titel „Milchproduktion pro Kuh pro Jahr“ vom der/dem Herausgebenden „Landwirt*in X“ sein, der/die an die Verwendungsbedingung (= Usage Policy) „Daten dürfen nur in der EU gespeichert werden“ und der Zugriffsbedingung (= Access Policy) „Mitglied im BIO-Laden-Klub“ angeboten wird. Wenn ein/eine Interessent*in diesen Bedingungen zustimmt, dann wird die technologische Grundlage zum Datentransfer etabliert (= Konnektoren verbinden sich).



Betrachtet man das auf einer oberflächlichen technologischen Ebene, bedeutet das, dass jeder/jede Teilnehmende seine Daten in seinen Datenquellen behält und immer nur Zugang zu den MetaDaten und den Policies gewährt. Erst, wenn den Policies zugestimmt wurde, erfolgt ein Datentransfer via Konnektoren. Ein Konnektor ist somit ein technisches Bauteil, das auf der einen Seite mit einer Datenquelle verbunden ist und auf der anderen Seite sich zu anderen Konnektoren zu definierten Bedingungen verbinden kann. Diese Konnektoren sind somit mehr als nur Adapter – vielmehr können sie zum Teil auch selbst sicherstellen, dass die Policies erfüllt werden. Beispielsweise könnte ein Konnektor die Information tragen, dass die daran angebundene Datenquelle in der EU liegt. Wäre dem nicht so, dann könnte Policies, die dies bedingen, gar nicht erst zugestimmt werden und es würde kein Transfer erfolgen (GAIA-X Data Exchange Service).

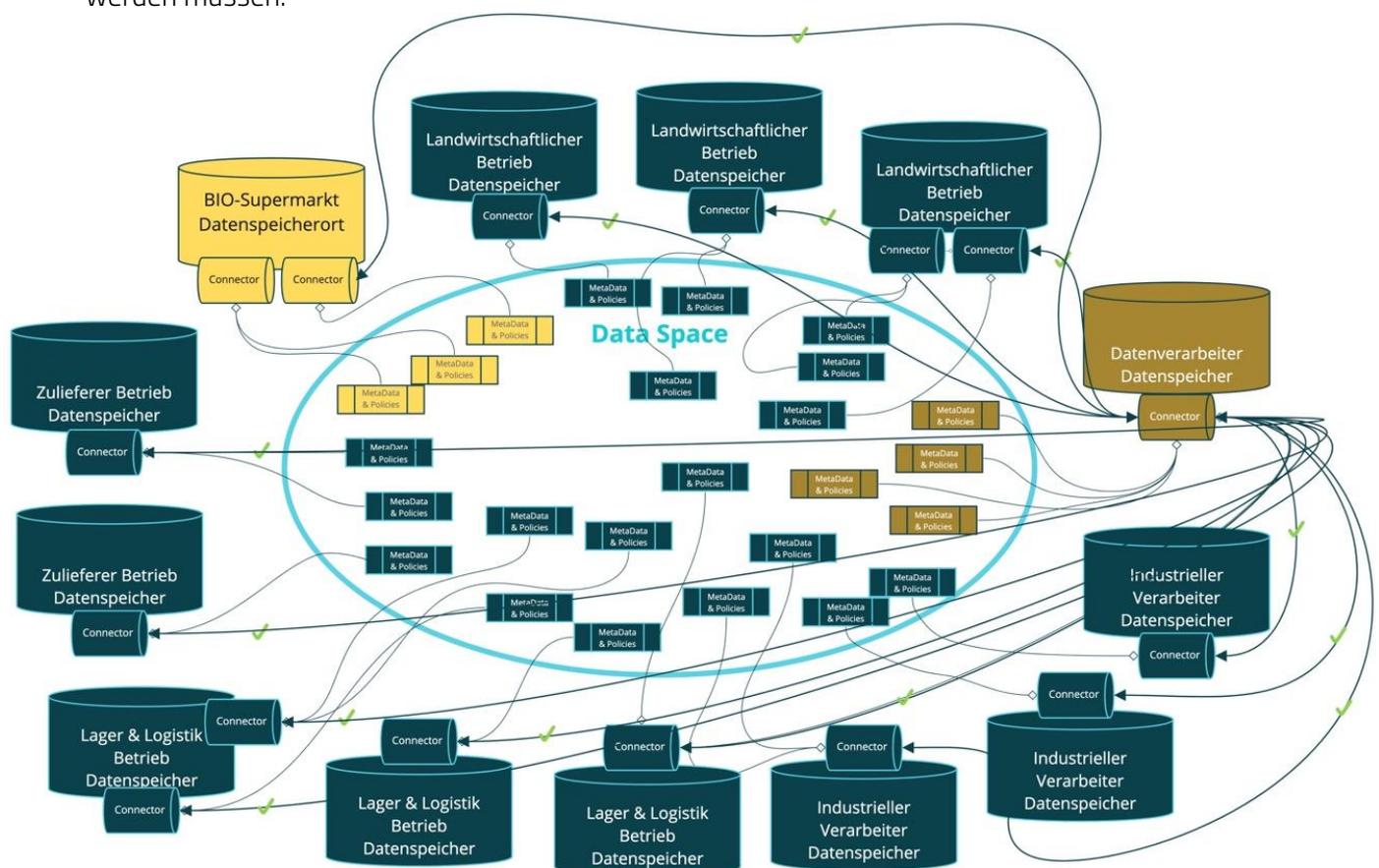


Zusammengefasst:

Die Notwendigkeit von Data Spaces wird schnell ersichtlich, wenn es darum geht Daten in einer vertrauensvollen Umgebung mit anderen außerhalb des eigenen Unternehmens austauschen zu wollen. In einer Welt, die immer stärker datengetrieben agiert und Produkte aus Daten kreiert, ist es unabdingbar auch technologisch die geeignete Umgebung dafür zu schaffen. Mit der Möglichkeit in neutralen, dezentralen digitalen Räumen, also Data Spaces, die in keinem Hoheitsgebiet einer einzelnen Partei liegen, wird die Grundlage für den Austausch von MetaDaten und die Einigung auf Policies geschaffen. Technologisch betrachtet bilden sogenannte Konnektoren die Basis dafür.

User Story Teil II

Alex hat es geschafft alle Zulieferer der Produkte in der Eigenmarken-Produktlinie zusammenzuführen. Das Bild, das sich ergibt, ist alleine für diesen Use Case schon ein sehr beachtliches! Zahlreiche Lieferant*innen und Logistiker*innen liefern zu ihren eigenen Bedingungen Daten, mit dem Ziel, den CO₂-Fußabdruck zu berechnen und auf der Verpackung abzubilden. Es werden also nicht nur Daten von den einzelnen Quellen geliefert, sondern sie werden auch von einem professionellen Unternehmen zur Datenanalyse verarbeitet und dann wieder zurückgespielt, damit der individuelle Wert auf jedes Etikett kommt. Durch die Vielzahl der Teilnehmenden ist eine manuelle Verwaltung der Datentransfers schon lange nicht mehr machbar. Selbst eine vereinfachte Darstellung zeigt schnell auf wie viele Verbindungen hier unter den ermittelten Auflagen etabliert werden müssen:



Das Projekt ist aber nicht nur Unternehmensintern ein Erfolg. Die erfolgreiche Umsetzung eines datenbasierten Produkts mit so vielen unterschiedlichen Stakeholdern, mit dem Ziel der Transparenz im Sinne der Nachhaltigkeit, zieht auch mediale Blicke auf sich. Diverse Zeitschriften, Blogs und Onlinemagazine greifen den Use Case auf und das Thema gewinnt weiter an Fahrt. Innerhalb von kürzester Zeit macht sich das auch in den Umsatzzahlen bemerkbar, da die

umweltbewussten Konsument*innen der BIO-Supermarktkette die Transparenz zu schätzen wissen.

Die gesteigerte mediale Aufmerksamkeit und die höheren Umsätze führen dazu, dass Alex von ihrer Vorgesetzten erneut einen Projektauftrag erhält. Sie soll herausfinden „Was sich da sonst noch so machen lässt, mit diesen Daten. Das scheint ja ganz gut zu funktionieren!“

Einer der Logistikpartner hatte bereits während des Proof of Concept angedeutet, dass er Interesse an einer datenbasierenden Routen-Optimierung hätte.

Also machen sich Alex, der Logistikpartner und das Datenverarbeitungs-Start Up dran einen weiteren Use Case zu formulieren, mit dem Ziel den CO₂-Ausstoß nicht nur zu messen, sondern auch zu optimieren. Die drei Parteien einigen sich darauf für die Vertragslaufzeit von einem Jahr Berechnungen zur Routenoptimierung anzustellen: auf Basis der weiter fortlaufend gesammelten Daten soll die Routenführung optimiert, Synergien zu anderen Lieferant*innen geschaffen und der Fortschritt gemessen werden. Nach Ablauf des Jahres sollen einerseits die automatisierten Datentransfers wieder eingestellt werden und andererseits, der optimierte CO₂-Ausstoß als Referenz für weitere Maßnahmen dienen.

Ersteres stellt kein Problem dar, da bereits auf technologischer Ebene durch die Konnektoren sichergestellt ist, dass ein Datentransfer nur zu den definierten Bedingungen erfolgen darf. Läuft also ein Vertrag aus, oder hält sich einer der Partner*innen nicht an eine der anderen Bedingungen, dann findet auch kein Datenaustausch statt, bzw. hat auch jeder/jede Datenlieferant*in zu jedem Zeitpunkt die volle Kontrolle über seine/ihre Transfers, was wiederum den CDO und die Rechtsabteilung der BIO-Supermarktkette begeistert.

Konnektoren bieten somit mehrere Vorteile auf einen Blick:

- Datentransfers erfolgen nur zu selbst definierten Bedingungen, denen ausdrücklich zugestimmt werden muss.
- Diese Bedingungen können so weitreichend sein, dass bereits auf technischer Ebene sichergestellt werden kann, ob der/die Datenkonsument*in überhaupt berechtigt ist, darauf zuzugreifen. Beispielsweise kann bereits auf dieser Ebene definiert werden, dass nur Mitglieder einer bestimmten Gruppe darauf zugreifen dürfen. (= Verifiable Credentials)
- Auf Konnektoren-Ebene kann auch bereits sichergestellt werden, dass der Ziel-Konnektor mit dem geplanten Speicherort geografisch auch den Richtlinien entspricht. Wenn also beispielsweise Daten nur in der EU gespeichert werden dürfen, ist auch das eine Information, die auf Konnektoren-Ebene sichergestellt werden kann.
- Transfers können auf mehrere Arten beschränkt werden, wie beispielsweise „darf nur 1x erfolgen“ oder „laufende Aktualisierung aus dem Produktionsbetrieb für eine definierte Dauer“. Außerhalb der definierten Bedingungen ist kein Zugriff darauf möglich.

Die Technologie und die Use Cases eröffnen den Teilnehmenden viele weitere Möglichkeiten und erregen auch weiter Aufmerksamkeit damit.

Von einem einfachen Proof of Concept, über ein datengetriebenes Produkt, das Transparenz schafft – dem CO₂-Fußabdruck-Label auf einer ganzen Produktpalette – bis hin zur Optimierung des tatsächlichen CO₂-Ausstoßes dieser Produkte. Alex hat nicht nur tatsächlichen Mehrwert für ihr Unternehmen, dessen Kund*innen und für die Umwelt geschaffen, sondern auch das Interesse vieler anderer geweckt!

Es ist nun nicht mehr nur interessant, wie einzelne Logistikpartner ihre Lieferwege optimieren und den CO₂-Ausstoß verringern können, sondern das Bestreben die ganze Supply-Chain bis hin zu den Düngemittel-Lieferant*innen und deren Zulieferern abzubilden und zu optimieren liegt nahe. Zusätzlich muss ob des Lieferketten-Gesetzes auch sichergestellt sein, dass alle Partner*innen definierten Anforderungen entsprechen, ethischen Standards genügen und integer sind, weswegen weitere Service Provider sich dem Data Space anschließen wollen.

Andere wiederum haben nur von dem Data Space gehört und sind bestrebt einen eigenen zu gründen, da auch gefälschter Gouda im B2B Bereich² ein ernsthaftes Problem darstellt und sich durch die neuen datengetriebenen technologischen Möglichkeiten auch neue Möglichkeiten ergeben, Problemstellungen wie diesen entgegenzuwirken.

Zusammengefasst kann man sagen, die Herausforderungen an Data Spaces sind:

- Viele unterschiedliche Parteien wollen gemeinsam datenbasiert Mehrwert erzeugen, sei es in datenbasierten Produkten, Services, Prozessoptimierungen oder Einsparungen
- Diese unterschiedlichen Parteien kennen sich aber untereinander nicht, oder sind unter Umständen sogar Konkurrent*innen
- Außerdem muss jede dieser Parteien zu jedem Zeitpunkt die Kontrolle über die eigenen Datenflüsse behalten und wissen wer wann auf die eigenen Daten zugreift
- Und zu welchen Bedingungen auf die Daten zugegriffen wird

All das lässt sich bereits mit bestehenden Technologien abbilden und verwirklichen, in dezentralen Systemen, die mittels Konnektoren verbunden werden und sich in Data Spaces treffen, um Zusammenarbeit zu ermöglichen.

² Vgl. <https://www.tagesspiegel.de/wirtschaft/verbrauchertauschung-gefaelschter-kaese-auf-der-pizza/1508850.html>

Data Spaces beschränken sich nicht auf Supply Chains, oder einzelne Use Cases im Nachhaltigkeitsbereich. Data Spaces können ganze Domänen vereinen wie Mobilität, Gesundheit oder Landwirtschaft. Data Spaces können aus der Wirtschaft, aber genauso aus der Forschung, oder politisch getrieben sein, oder sogar auch eine Kombination aus allen.

Es geht um Souveränität, Trust und Innovation in dezentralen Ökosystemen, all das und noch viel mehr ist mittlerweile technologisch möglich.

Wir müssen es nur noch tun.