



Mittelstand-Digital
Zentrum
WertNetzWerke



© KI-generiertes Bild

Leitfaden

Sicherer Datenaustausch in Wertschöpfungsnetzwerken

Gefördert durch:



Bundesministerium
für Wirtschaft
und Klimaschutz

Mittelstand-
Digital 

aufgrund eines Beschlusses
des Deutschen Bundestages

Inhaltsverzeichnis

Einleitung	3
Die Rolle von Wertschöpfungsnetzwerken in der Produktion	3
GAIA-X	4
Datenräume	4
Data Sovereignty as a Service	4
Catena-X	4
Manufacturing-X	5
Factory-X	5
Weitere Ansätze und Lösungen für den vertrauensvollen Datenaustausch	5
Confidential Computing	5
Mehr-Parteien-Computing	6
Industrie-Clouds	9
Datenmarktplätze	9
Datengenossenschaften	10
Ausblick	10
Weitere Informationen/Quellenangaben	11
Impressum	12

Einleitung

Wenn es um den Datenaustausch mit Dritten geht, ist die Zurückhaltung in den kleinen und mittleren Unternehmen (KMU) noch stark ausgeprägt. Die Sorge ist groß, dass sensible Daten in die falschen Hände geraten oder zweckentfremdet werden könnten.

Aktuell werden die Risiken daher als zu hoch eingeschätzt. Dabei ist es gar nicht mal so sehr die Frage, ob die Daten sicher von A nach B übertragen werden können. Hier ist das Vertrauen in die Verschlüsselungstechnologien durchaus vorhanden. Schwerer wiegt dagegen, wenn sich sensible Maschinen-, Prozess-, Qualitäts- und Energiedaten außerhalb der eigenen Kontrolle bei Dritten befinden. Stand heute kann nicht ausgeschlossen werden, dass Dritte auf irgendeine Weise in der Lage sind, die Anwendungsfälle zu rekonstruieren.

Andererseits erfordert die Zusammenarbeit in Wertschöpfungsnetzwerken ein Mindestmaß an der Bereitstellung sensibler Daten. Abgesehen davon setzt die Entwicklung datenbasierter Geschäftsmodelle in vielen Fällen eine bestimmte Menge an Daten für das Training von KI-Modellen voraus. Die meisten KMU verfügen jedoch nicht über die dazu nötige Datenbasis. Damit besteht die Gefahr, dass die KMU bei der Digitalisierung ihrer Geschäftsmodelle von den großen Unternehmen abgehängt werden, die über die nötigen technischen, personellen und organisatorischen Kapazitäten verfügen. Ein Weg könnte der Zusammenschluss mehrerer KMU zu einer Datenkooperation oder Datengenossenschaft sein.

Mittlerweile stehen einige Verfahren und Technologien zur Verfügung, die sich jedoch noch im Anfangsstadium befinden. Am bekanntesten dürfte hierzulande GAIA-X sein, die europäische Dateninfrastruktur, welche die Datensouveränität der Unternehmen sicherstellen soll. Deren primäres Ziel ist es, die Abhängigkeit der europäischen Wirtschaft von den sog. Hyperscalern, wie Amazon, Microsoft und Google, zu verringern. Gemeinsame Datenräume auf Basis von GAIA-X könnten KMU neben dem sicheren Austausch der Daten auch den Zugang zu großen Datenpools verschaffen.

Auf dem Gebiet neuer Technologien sind Confidential Computing, Multi Party Computation, Federated Learning und der Einsatz synthetischer Daten zum gegenwärtigen Zeitpunkt vielversprechend. Der Vorteil dieser Lösungen besteht darin, dass

keine realen Daten übertragen oder für Auswertungszwecke verwendet werden.

Unter den kommerziellen Anwendungen stechen derzeit die Industrie-Clouds der Hyperscaler hervor.

Die KMU stehen vor der schwierigen Frage, auf welche Lösung und welchen Trend sie setzen sollen, was angesichts der Vielzahl der zur Wahl stehenden Optionen kaum zu bewältigen ist.

Der vorliegende Report dient dem Zweck, den aktuellen Stand an verfügbaren Lösungen für den sicheren Datenaustausch in Wertschöpfungsnetzwerken zu veranschaulichen, um dadurch vor allem KMU einen ersten bzw. besseren Überblick über die Thematik zu verschaffen.

Die Rolle von Wertschöpfungsnetzwerken in der Produktion

Die Leistungserstellung vollzieht sich heutzutage immer häufiger über die Unternehmensgrenzen hinweg. Vorbei die Zeiten, als ein Unternehmen erst dann mit der Außenwelt kommunizieren musste, nachdem die Leistung erbracht bzw. das Produkt erstellt war. Heute müssen die Unternehmen im ständigen Austausch stehen, um z. B. kostspielige Verzögerungen in der Lieferkette zu verhindern. Beispielhaft dafür sind die Just-in-time-Fertigung in der Automobilindustrie und die vorausschauende Wartung im Maschinenbau. Ein Produkt bedarf auch noch lange Zeit nach der Auslieferung der Betreuung durch den Hersteller. Mehr noch: In der Produktion muss die Möglichkeit bestehen, neue Anforderungen der Kunden zu berücksichtigen. Ohne unternehmensübergreifenden, multilateralen Datenaustausch, wie z. B. zwischen Maschinenhersteller, Maschinenbetreiber, Systemintegrator und Komponentenhersteller fehlt der vernetzten Wertschöpfung die Basis. Das Potenzial der Daten bleibt dann ungenutzt:

„Während der Wertschöpfungsfluss physischer Produkte hauptsächlich unidirektional entlang der Lieferkette von den Lieferanten über das Unternehmen zur Kundschaft verläuft, liegt das Nutzenpotenzial von Daten zumeist in einem bidirektionalen und oftmals multilateralen Austausch. Um das Potenzial gänzlich zu erfassen, ist folglich eine Erweiterung des Betrachtungsrahmens vom Unternehmen auf das gesamte Ökosystem notwendig, in dem weitere Entitäten, wie Forschungseinrichtungen und Drittfirmen, agieren. Die interne Betrachtung des Wertschöpfungsflusses wird somit um

eine externe Perspektive ergänzt, die die Einbindung externer Daten, den Austausch zum gemeinsamen Nutzen sowie den Verkauf und Handel mit Daten ermöglicht.“¹

Kurzum: Das Produkt sendet bis zu seinem Lebensende Daten, die Auskunft über seinen Zustand, sein näheres Umfeld und seine Nutzung geben. Diese Daten lassen sich wiederum für die Verbesserung bestehender und die Erstellung neuer Produkte und Services verwenden.²

GAIA-X

Bei GAIA-X handelt es sich, anders als hin und wieder noch angenommen wird, um eine digitale Infrastruktur und keine Cloud-Lösung, wie sie von den sog. Hyperscalern angeboten wird. Sie stellt das technische, organisatorische und rechtliche Rahmenwerk bereit, anhand derer die Cloud-Lösungen genutzt werden können. Das Hauptaugenmerk liegt auf dem sicheren und souveränen Datenaustausch. Die Unternehmen sollen selbst bestimmen können, wer auf welche Daten, für welchen Zweck und für welchen Zeitraum zugreifen darf. Nachdem die Anforderungen der KMU ein wenig unter den Tisch fielen, sind sie in letzter Zeit verstärkt in den Fokus gerückt. In verschiedenen Projekten wird versucht, die hoch komplexe Referenzarchitektur an die Bedürfnisse kleiner und mittlerer Unternehmen anzupassen.

Datenräume

Wenn Unternehmen davor zurückschrecken, Daten direkt mit Dritten zu tauschen, besteht die Möglichkeit, die Daten in einem separaten, geschützten Raum (Data Space) hochzuladen, wo sie dann von den dazu berechtigten Unternehmen oder Organisationen abgerufen werden können. Am bekanntesten sind die verschiedenen Datenräume, die im Rahmen von GAIA-X angeboten werden, wie Catena-X, Manufacturing-X und der Mobility Data Space. Ziel ist es, für jede Branche einen Datenraum zur Verfügung zu stellen.³

„Ein Datenraum ermöglicht den Austausch von Daten – dies ist seine Kernfunktion. Teilfunktionen eines Datenraumes sind Funktionen, die die Bereitstellung, den Empfang und die Übermittlung von Daten (die logischen Bestandteile des Austausches) ermöglichen, erleichtern oder steuern. Hierzu zählen beispielsweise Funktionen zur Katalogisierung oder Identifikation von Teilnehmern und Gegenständen des Austauschs.“⁴

Lösungen, die GAIA-X – konform sind, verpflichten sich, die geltenden EU-Bestimmungen, wie die DSGVO, umzusetzen. Ferner bieten sie über den sog. Eclipse Data Connector (EDC) den Unternehmen die Möglichkeit, die Regeln für die Nutzung der Daten selbst zu bestimmen, d. h. für welchen Zweck und Zeitraum die Daten genutzt werden können.⁵

Mittlerweile ist man bei GAIA-X bestrebt, auch kleine und mittlere Unternehmen einzubeziehen. GAIA-X bietet mittlerweile einmal im Monat Onboarding-Webinare für Interessierte und Mitglieder an.⁶

Ein Beispiel für die Umsetzung der Idee der Datenräume in der Praxis ist die Regional Edge Cloud. Die Lösung stellt eine sichere Vernetzung, die sicheren Datenräume und eine Regional Edge Cloud zur Verfügung.⁷ Damit will man die sog. Shared Production ermöglichen. Der regionale Bezug rührt daher, da die Feldebene direkt per Glasfaser an lokale Rechenzentren angebunden wird.

Data sovereignty as a Service

Um KMU den Anschluss an GAIA-X – konforme Datenräume mit minimalem Aufwand zu ermöglichen, können sie die Dienste spezialisierter Anbieter nutzen. Dabei übernimmt der Dienstleister alle technischen Dinge. Das KMU erhält z. B. eine Mail mit einem Link, über den der Zugang zum Connector hergestellt wird. Dann kann das Unternehmen seine Datenquellen anbinden und über den jeweiligen Data Space Daten austauschen. Allerdings erfordert das auf Seiten des KMU Vorarbeiten, was die Bereitstellung der Daten in der nötigen Qualität betrifft. Eine weitere Herausforderung ist es, die eigenen Daten an die passenden Datenformate anzupassen.

Catena-X

Catena-X versteht sich als erweiterbares Ökosystem, an dem sich Automobilhersteller und -zulieferer, Händlerverbände sowie Ausrüster, zu denen Anwendungs-, Plattform- und Infrastrukturanbieter gehören, gleichermaßen beteiligen können. Ziel ist es, einen einheitlichen Standard für Daten- und Informationsflüsse in der gesamten automobilen Wertschöpfungskette zu schaffen. Neben Effizienzvorteilen in der Lieferkette versprechen sich die Teilnehmer des Netzwerkes beispielsweise leistungsfähigere Qualitäts- und Logistikprozesse, höhere Transparenz hinsichtlich nachhaltiger CO₂-

Reduzierung sowie ein vereinfachtes Stammdatenmanagement. Auf diese Weise ermöglichen es durchgehend verbundene Datenketten, digitale Zwillinge von Automobilen zu erschaffen, auf deren Basis sich innovative Geschäftsprozesse und Serviceangebote entwickeln lassen.

Catena-X richtet sich in erster Linie an die OEMs und deren Tier-1-Zulieferer. Sobald Catena-X für die Zulieferer verpflichtend wird, kommt das Thema über kurz oder lang auch bei den KMU an.

Getragen wird Catena-X von der Betreibergesellschaft Confinity-X. Gesellschafter sind BASF, BMW Group, Henkel, Mercedes-Benz, SAP, Schaeffler, Siemens, T-Systems, Volkswagen und ZF. Die zehn Gesellschafter sind zu gleichen Teilen an dem Joint Venture beteiligt.⁸ Der Mittelstand, geschweige denn KMU, sind darin nicht vertreten, was sich als Konzeptionsfehler erweisen könnte.

Manufacturing-X

Industrie 4.0 braucht eine einfach verfügbare, sichere und durchgängige Datenvernetzung. Manufacturing-X soll das ermöglichen. Ziel ist die Etablierung eines Daten-Ökosystems, das den vertrauensvollen, auf offenen Standards basierenden Datenaustausch zwischen Unternehmen ermöglicht sowie den Firmen digitale Souveränität bietet.

Manufacturing X soll neue Geschäftsmodelle entwickeln und die Wettbewerbsfähigkeit der Industrie durch die Nutzbarmachung von digitalen Technologien wie dem Internet der Dinge (IoT), künstlicher Intelligenz (KI) und maschinellem Lernen (ML) stärken. Durch den Einsatz dieser Technologien können Unternehmen ihre Produktionsprozesse optimieren, ihre Effizienz steigern und innovative Lösungen für ihre Kunden entwickeln.⁹

Im Unterschied zu Catena-X, wo die Unternehmen hauptsächlich Stammdaten „in der Zulieferkette austauschen, d. h. Werkstoffhersteller mit Teilefertigern, diese mit Baugruppenlieferanten, diese mit Modulherstellern, diese wieder mit den OEMs“, hat für die produzierende Industrie „der Austausch von Daten aus dem Betrieb von Komponenten, Maschinen und Anlagen“ Priorität.¹⁰

Um die Einstiegsbarrieren für KMU gering zu halten, müssen überzeugende Migrationskonzepte in den Datenraum für KMU erarbeitet werden. „Es müssen einfache Teilhabemöglichkeiten an Datenräumen geschaffen werden (z. B. ein erleichter-

ter Daten Upload Service). Die Anforderung einer partiellen Teilnahme an einem Datenraum muss berücksichtigt werden. So sollte es möglich sein, dass einzelne Anlagen oder Komponenten Teil des Datenraums sind, auch dann, wenn die Fabrik, in der die Anlage steht, noch nicht in einen Datenraum integriert ist.“¹¹

Factory-X

Derzeit planen SAP und Siemens gemeinsam mit rund 50 Partnern aus Wirtschaft, Forschung und Verbänden, eine digitale Plattform für den Datenaustausch in der Fertigungsindustrie aufzubauen.¹² Die Systeme werden dafür mit Datenraum-Konnektoren ausgestattet.

Weitere Ansätze und Lösungen für den vertrauensvollen Datenaustausch

Für den vertrauensvollen Datenaustausch stehen inzwischen mehrere technische Lösungen zur Verfügung, welche die modernsten kryptografischen Verfahren und die sicherste am Markt verfügbare Hardware einsetzen. Ein prominentes Beispiel ist das Confidential Computing. Neben dem Confidential Computing existieren weitere Techniken und Kombinationen von Ansätzen zur Sicherung von Daten im Einsatz, wie homomorphe Verschlüsselung (HME), sicheres Mehrparteienrechnen, Nullwissen und synthetische Daten. Die Unterschiede bestehen in der Hardware und der Art und Weise, wie sie in Verbindung mit Software eingesetzt wird. Nicht alle Anbieter verfügen über das gleiche Schutzniveau.

Confidential Computing

Heutzutage können Unternehmen der öffentlichen Cloud nicht vollständig vertrauen, da sie nicht sicher wissen, wer – wie Hacker oder böswillige Administratoren – auf ihre Daten zugreifen könnte. Die entscheidende Frage lautet daher: Wie lassen sich die Workloads isolieren und vor dem Zugriff (ausländischer) Infrastrukturanbieter schützen?

Beim Confidential Computing wird ein Teil des Prozessors und des Speichers gesichert, um eine geschützte und isolierte Umgebung für den Code und die Daten zu schaffen, die als vertrauenswürdige Ausführungsumgebung (Trusted Execution Environment) bezeichnet wird. Auf diese Weise werden Daten vor unbefugtem Zugriff oder Manipulationen geschützt.

Projekte auf diesem Gebiet u. a. Intel Software Guard Extension, Microsoft Open Enclave SDK und Amber.¹³ Daneben gibt es noch eine Reihe weiterer, kleinerer Anbieter - auch in Deutschland.

Confidential Computing ist jedoch kein Plug-and-Play. Die Interaktion mit sicheren Enklaven erfordert den Einsatz spezialisierter Technologien. Das größte Risiko besteht laut Brancheninsidern in der Implementierung. Je nachdem, wie man eine vertrauliche Computerumgebung strukturiert, verschlüsselt man im Grunde alle Daten, damit sie nicht in die falschen Hände geraten – aber man könnte sich auch selbst aussperren.

In naher Zukunft werden laut Marktbeobachtern alle Workloads mit Confidential Computing implementiert werden, um die Sicherheit zu gewährleisten. Allerdings werde es Zeit brauchen, bis vorgelagerte Schichten wie Linux-Distributoren es integrieren, und noch mehr Zeit, bis ein Ökosystem von Anbietern es nutzt.

Aufgrund der derzeit noch vorhandenen Risiken und des nicht zu unterschätzenden Aufwands ist das Confidential Computing für die meisten KMU noch keine echte Alternative. Man sollte die Entwicklung jedoch im Auge behalten.

Mehr-Parteien-Computing

Die sichere Mehrparteienberechnung (auch bekannt als sichere Berechnung, MPC oder datenschutzfreundliche Berechnung) ist ein Teilgebiet der Kryptographie, das darauf abzielt, Methoden zu entwickeln, mit denen Parteien gemeinsam eine Funktion über ihre Eingaben berechnen können, wobei diese Eingaben geheim bleiben. Im Gegensatz zu herkömmlichen kryptografischen Aufgaben, bei denen die Kryptografie die Sicherheit und Integrität der Kommunikation oder der Speicherung gewährleistet und der Gegner außerhalb des Systems der Teilnehmer steht (ein Lauscher beim Sender und Empfänger), schützt die Kryptografie in diesem Modell die Privatsphäre der Teilnehmer voneinander.¹⁴

Das ist eine große Errungenschaft, da Unternehmen oft befürchten, dass das Risiko, ihre Daten offen zu teilen, größer ist als der potenzielle Wert der gemeinsamen Nutzung selbst. Ein Beispiel ist die Lieferkette, an der mehrere Parteien beteiligt sind. Die Verfolgung auf Objektebene ermöglicht es Unternehmen, große Datenmengen zu sammeln, z. B. über die Zeit, den Ort oder die Handhabung

der von ihnen produzierten Waren. Die Zusammenführung der von den verschiedenen beteiligten Unternehmen gesammelten Daten kann erhebliche Vorteile mit sich bringen. Es kann jedoch sein, dass zwei Unternehmen nur bereit sind, Informationen über gemeinsame Gegenstände auszutauschen, die sie beide entlang der Lieferkette bearbeitet haben. MPC kann dieses Dilemma mit sicheren und privaten Berechnungen lösen.

Derzeit ist die MPC für KMU eher ein perspektivisches Thema, über das man sich auf dem Laufenden halten sollte.

Homomorphe Verschlüsselung

Die Homomorphe Verschlüsselung gestattet es, Berechnungen mit Daten durchzuführen, ohne diese vorher entschlüsseln zu müssen. Die Qualität der Ergebnisse ist so gut, als wären sie mit unverschlüsselten Daten erfolgt.¹⁵

Im Projekt [SMiLe](#) wird untersucht, unter welchen Voraussetzungen Lösungen, bei denen homomorphe Verschlüsselung zum Einsatz kommen, dazu geeignet sind, das Potenzial von sensiblen Daten für maschinelles Lernen nutzbar zu machen. Die Weiterentwicklung von relevanten Softwarekomponenten steht genauso im Mittelpunkt des Projekts wie das Zugänglichmachen von erforderlichem Know-how. Das Potenzial des Verfahrens für maschinelles Lernen wird anhand von zwei Anwendungsfällen bewertet, die sich mit Mitarbeitersegmentierung bzw. vorausschauender Wartung befassen. Im Rahmen des Projekts werden sowohl technische als auch soziale, rechtliche und wirtschaftliche Fragen behandelt. Lösungsansätze werden nicht nur hinsichtlich ihrer Analysefähigkeiten, Leistung und Skalierbarkeit, sondern auch im Hinblick auf ihre Kosteneffizienz, Transparenz und Benutzerfreundlichkeit bewertet. Die Vor- und Nachteile des maschinellen Lernens auf homomorph verschlüsselten Daten werden mit alternativen Ansätzen verglichen, die auf synthetischen Daten, Transfer Learning, Secure Multi-Party Computation oder Differential Privacy basieren.

Nachteilig bei der homomorphen Entschlüsselung ist, dass sie einen hohen Rechenbedarf hat und langsam ist.¹⁶

Föderiertes Lernen (FL)

Die Zusammenführung von Trainingsdaten über viele Unternehmen hinweg würde auch kleine und mittlere Unternehmen in die Lage versetzen, maschinelle Lernsysteme für die vorausschauende

Wartung einzusetzen. Nötig wäre ein kollaboratives Modell für die vorausschauende Wartung unter Nutzung des unternehmensübergreifenden Datenaustauschs, ohne dabei die Geschäftsinformationen der Teilnehmer preiszugeben.

Das Teilen von Daten zwischen mehreren KMU oder Maschinenherstellern, um ein leistungsfähiges Modell für die vorausschauende Wartung zu trainieren, funktioniert nur, wenn Datenintegrität und Datensicherheit gewährleistet sind. Werden Daten mit externen Partnern ausgetauscht, könnte ein böswilliger Akteur deren Daten verfälschen, um das Modell zu sabotieren oder die Modellvorhersagen subtil zu seinen Gunsten zu verändern. Ein Ansatz diese Risiken zu umgehen, zumindest aber zu reduzieren ist, statt der Daten nur die KI-Modelle zu tauschen: Föderiertes Lernen.

Ein FL-Algorithmus arbeitet in mehreren Runden. Zu Beginn erstellt der zentrale Kurator eine Liste der teilnehmenden Maschinen. Danach werden mithilfe eines Kandidatenauswahlalgorithmus kooperative Teilnehmer mit geeigneten Daten ausgewählt. Das aktuelle globale Modell wird an die zuständigen Mitarbeiter der Unternehmen gesen-

det. Diese verwenden dann ihre privaten Daten unter Verwendung ihrer eigenen Hardware, um ein aktualisiertes lokales Modell abzuleiten und es an den Kurator zurückzusenden. Die lokalen Modelle werden vom Kurator in mehreren Runden zusammengefasst, bis es genau genug ist. Danach verteilt der Kurator das resultierende Modell an die Teilnehmer.

Nachfolgend werden einige Projekte vorgestellt, die sich damit beschäftigen, die Möglichkeiten des föderierten Lernens in der Praxis zu erforschen und noch bestehende Defizite zu beheben.

KOSMoS

Jeder Maschinenbetreiber, der an dem Projekt teilnehmen möchte, benötigt ein Edge Device lokal auf seinem Hallenboden, das in den privaten Teil der Architektur von [KOSMoS](#) integriert ist. Das Edge Device hat Zugriff auf alle Maschinendaten, die für das Modelltraining benötigt werden. Zusätzlich dient es als Kommunikationsschnittstelle. Unterschiedliche Datenstrukturen der Maschinen und Sensoren werden in Topic-Nachrichten übersetzt, sodass die Modelle mit einer einheitlichen Struktur trainiert werden können.

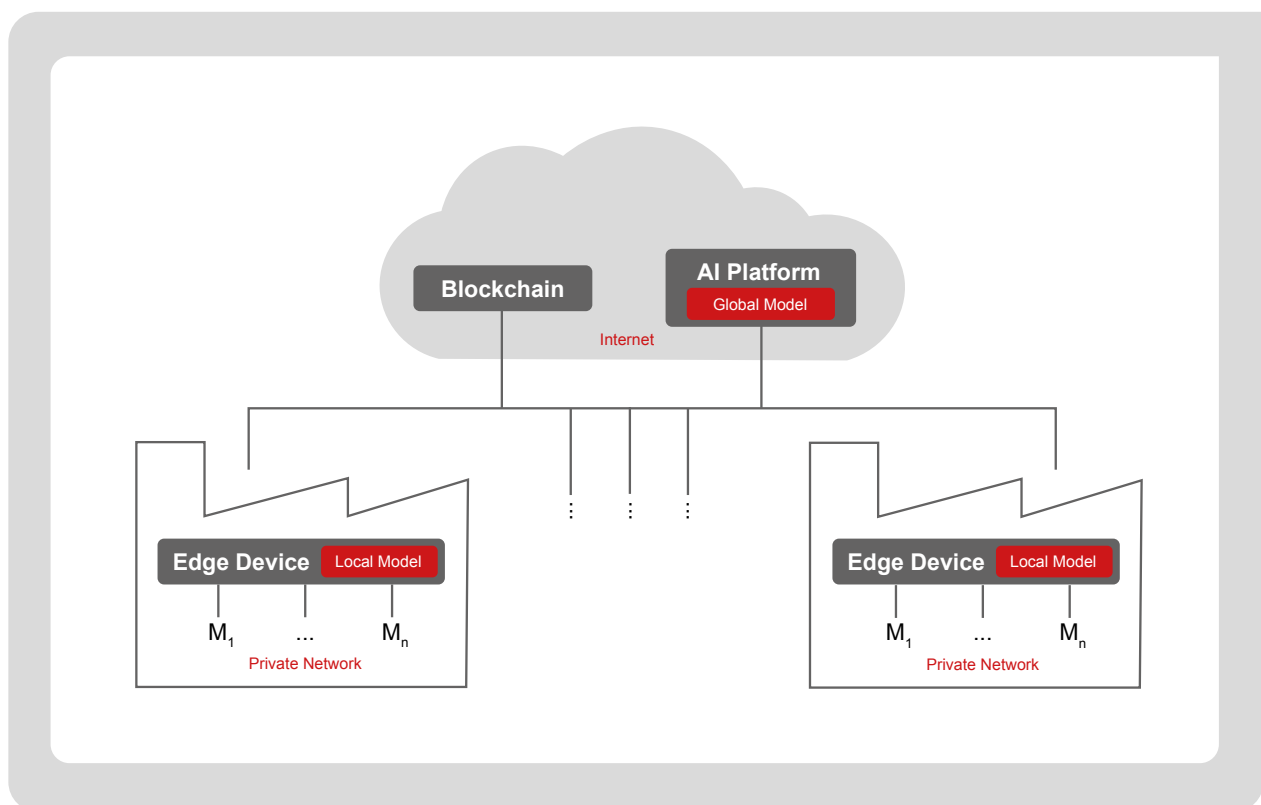


Abbildung 1: Federated Learning in KOSMoS

Maschinenhersteller, Maschinennutzer und auch Dritte wie Servicepartner können mittels gefördertem Lernen und der Blockchain Wertschöpfungsketten zu Wertschöpfungsnetzwerken erweitern und damit datengesteuerte Geschäftsmodelle schaffen, die über Unternehmensgrenzen hinweg genutzt werden.

FedZero

Die wachsende Nutzung von Federated-Learning-Anwendungen erhöht den weltweiten Energieverbrauch des maschinellen Lernens signifikant. Maschinelles Lernen gilt ohnehin als eine der energieintensivsten Rechenanwendungen. Das Konzept des verteilten Trainings von KI-Modellen in örtlich getrennten Datenzentren bietet die Chance, die Ausführung der Rechenanwendungen an die Verfügbarkeit von grüner Energie zu koppeln. [FedZero](#) ist so angelegt, dass alle Trainingsprozesse nur ablaufen, wenn ausreichend grüne Energie zur Verfügung steht.

FLInK

Maschinelles Lernen (ML) bietet in der fertigen Industrie wirtschaftliche und ökologische Vorteile durch die datengetriebene Optimierung von Anlagenprozessen. Für Unternehmen mit wechselnden Prozessanforderungen in der Produktion ist ML jedoch herausfordernd, da umfangreiche Daten und Zeit für die Implementierung spezifischer KI-Modelle benötigt werden. Im Verbundprojekt [FLInK](#) wird daher untersucht, wie sich verteilte Lernverfahren und Schwarmintelligenz zum beschleunigten Transfer vortrainierter KI-Modelle in der fertigen Industrie nutzen lassen.

Innerhalb der Unternehmen findet verteiltes Lernen über eine zentrale Recheneinheit und Modelldatenbank statt. Der Transfer auf neue Anlagen kann durch die Verwendung bestehender Modelle zügig erfolgen. Der Einsatz eines dezentralen Schwarmnetzwerkes ermöglicht zusätzlich unternehmensübergreifendes verteiltes Lernen durch den Austausch spezifischer Modell-Parameter. *„Die Rohdaten bleiben dabei innerhalb des Unternehmens, wodurch ein Maximum an Datensicherheit und -anonymität erreicht wird“¹⁷*, so Projektmitarbeiter Dr. Michael Haub. Das verteilte Training und der Transfer generischer Modelle ermöglicht dadurch Unternehmen, ML mit erheblicher Zeitersparnis maßgeschneidert für neue Produkte oder Anlagen zu nutzen. Ein Punkt der vor allem für KMU, die für gewöhnlich nicht über die nötige Menge an Trainingsdaten verfügen, von Bedeutung ist.

Factory-X

Das bereits erwähnte Forschungsprojekt Factory-X will ebenfalls vortrainierte KI-Modelle integrieren, welche sich durch förderiertes Lernen weiterentwickeln.

Allerdings kommt auch dieser Ansatz nicht ohne Risiken daher. So können beim Federated Learning Modelle prinzipiell über manipulierte Trainingsdaten kompromittiert oder mit Hintertüren, sogenannten Backdoors, versehen werden. Ein vertrauenswürdiges KI-Modell auf Basis von Federated Learning sollte daher in der gesamten Lieferkette auf Risiken überprüfbar sein, idealerweise mit Methoden für Risikoanalyse und zur Risikominimierung im operativen Machine Learning-Prozess.

Der Aufwand für die Gewährleistung der Sicherheit und der Datenintegrität ist daher nicht zu unterschätzen. Vortrainierte KI-Modelle, die auf die Anforderungen von KMU angepasst sind, wären ein wichtiger Schritt nach vorne. Der Knackpunkt ist und bleibt jedoch – neben Frage der IT-Sicherheit – die ausreichende Menge und Qualität der Trainingsdaten.

Ein Weg, dieses Dilemma zu lösen, könnten Synthetische Daten sein.

Synthetische Daten

Synthetische Daten werden künstlich produziert und nicht durch reale Ereignisse generiert. Sie werden mittels Algorithmen erstellt und als Ersatz für Testdatensätze von Produktions- oder Betriebsdaten verwendet, um mathematische Modelle zu validieren und Modelle (Machine Learning, ML) zu trainieren.

Für die Gewinnung synthetischer Daten wird ein Modell entwickelt, das die Originaldaten so gut wie möglich erklärt. Einige Experten behaupten sogar, dass synthetische Daten für das Training von KI-Modellen besser sind als Daten realer Menschen, Orte und Dinge. Einschränkungen bei der Verwendung sensibler und regulierter Daten entfallen oder werden reduziert; Datensätze können auf bestimmte Bedingungen zugeschnitten werden, die sonst vielleicht nicht zu erreichen wären; Erkenntnisse können viel schneller gewonnen werden; und das Training ist weniger mühsam und viel effektiver.

Dateninnovation, so die Studie [Synthetische Daten Innovationspotential und gesellschaftliche Herausforderungen](#), „wird nur möglich sein, wenn wir

robuste Verfahren finden, dieses Risiko (das inhärente Risiko jeder Datennutzung) zu reduzieren. Daten-Synthetisierung könnte ein solches Verfahren sein, das dann zur vollen Entfaltung kommt, wenn die wichtigsten rechtlichen und ethischen Fragen geklärt sind“.

Allerdings sollten die Risiken, wie in Gestalt manipulierter oder verzerrter Daten, nicht außer Acht gelassen werden.¹⁸

Bei Gartner geht man übrigens davon aus, dass synthetische Daten bis 2030 reale Daten in KI-Modellen vollständig verdrängen werden.¹⁹ Für KMU könnte der Einsatz Synthetischer Daten daher ein sinnvoller Weg sein, um ausreichend Daten für das Training eigener Modelle zu generieren, ohne dabei Gefahr zu laufen, sensible Informationen nach außen, in die falschen Hände zu geben.

Industrie-Clouds

Seit einiger Zeit nimmt die Zahl der Industrie-Clouds zu. Branchenbeobachter gehen davon aus, dass Industrie-Clouds in einigen Jahren die bevorzugten Unternehmens-Applikationen sein werden. Tech-nopedia definiert die Industrie-Cloud als ein Cloud-System, das stark an die Bedürfnisse einer bestimmten Branche angepasst wurde, um den geschäftlichen, betrieblichen, rechtlichen, regulatorischen, sicherheitstechnischen und anderen Überlegungen Rechnung zu tragen. Das Hauptaugenmerk der Industrie-Cloud liegt auf der vertikalen Integration und nicht auf der horizontalen Integration, die beim allgemeinen Cloud Computing im Vordergrund steht.²⁰

Mittlerweile bieten Microsoft, SAP und Salesforce entsprechende Lösungen an. Volkswagen hat derweil seine Industrie Cloud für seine Partner aus dem Maschinenbau und Technologiesektor geöffnet.

Momentan arbeiten Volkswagen und AWS an einem App Store, in dem die Beteiligten ihre Anwendungen untereinander – und unabhängig von einem Einsatz in Volkswagen Werken – tauschen, erwerben und nutzen können.

Ein weiteres Beispiel ist Skywise von Airbus. Skywise ist eine Enterprise-Data-Plattform für die Luftfahrt-industrie. Sie umfasst über 100 Fluggesellschaften jeder Größe und Geographie. Die Plattform ist konzipiert für die Integrationen von kommerziellen und operativen Systemen und die Verarbeitung

großer Datenmengen, sowohl strukturierte (Betriebs- und Wartungsdaten) und unstrukturierte Daten (technische Dokumente). Es bietet den Benutzern Werkzeuge zur Vorbereitung, Aggregation und der Analyse von Daten sowie Vorlagen für die Erstellung von Anwendungen innerhalb der Plattform. Ein Wermutstropfen ist allerdings, dass die Plattform zusammen mit Palantir entwickelt wurde.

Für KMU derzeit wohl keine echte Alternative, da Industrie-Clouds vorwiegend an den Bedürfnissen von Großunternehmen ausgerichtet sind.

Datenmarktplätze

In den letzten Jahren sind einige B2B-Datenmarktplätze entstanden. Sie stellen die Infrastruktur für den Datenaustausch bereit, wobei sie als digitaler Vermittler („Intermediär“) eine Verbindung zwischen Datenanbieter und Dateneinkäufer schaffen und Mehrwertdienste, z. B. KI-gestützte Datenanalysen, anbieten. Die größten Datenerzeuger sind Maschinen und andere technische Geräte, weshalb Marktplätze für IoT-Daten an Zahl deutlich zugenommen haben. Ein Datenmarktplatz kann von der Organisation her proprietär/geschlossen sein, von einem Konsortium betrieben werden oder neutral/unabhängig auftreten.

Aus technischer Perspektive können Datenmarktplätze zentral (Cloud) oder dezentral (Edge/Cloud) konstruiert sein. Bei der Cloud-Variante werden die Datenprodukte von verschiedenen Anbietern über eine zentrale Stelle angeboten. Das hat den Vorteil, dass die Daten durch die Verwendung einheitlicher Datenmodelle- und Formate leicht durchsucht werden können. Nachteilig für die Datenerlieferanten ist der Verlust der Datenhoheit. Das Nutzungsrecht an den Daten liegt in den meisten Fällen bei dem Betreiber des Marktplatzes, der die Geschäftsbedingungen und technischen Zugangsvoraussetzungen festlegt.

Beim dezentralen Ansatz hingegen behält der Lieferant die Hoheit über seine Daten. Der Austausch erfolgt bilateral. Die Daten werden an die Nutzungsbedingungen des Lieferanten gekoppelt. Allerdings wird der durchgängige Datenaustausch durch die hohe Anzahl unterschiedlicher Interfaces, Formate und Preismodelle eingeschränkt.

Im Projekt [DIONE-X](#) wird derzeit das Zusammenspiel von Datenmarktplätzen und vortrainierten Modellen erprobt.

Für KMU kann der Einsatz von Datenmarktplätzen in bestimmten, überschaubaren Anwendungsfällen sinnvoll sein.

Datengenossenschaften

Eine Datengenossenschaft in der Rechtsform einer eG übernimmt die gemeinsame Sammlung und Analyse von Daten in einem Unternehmensnetzwerk bzw. in einem Ökosystem. Die Genossenschaftsmitglieder sind Anteilseigner der Genossenschaft. Sie legen die Regeln für die Organisation der Genossenschaft und für die Verfahren des Datensammelns, der Datenanalyse, Datenbewertung und Datenmonetarisierung fest. Die Datengenossenschaft stellt sie die nötige technische Infrastruktur und das Personal zur Verfügung. Für KMU kann die Mitgliedschaft in einer Datengenossenschaft eine ökonomisch sinnvolle Entscheidung sein.

Laut der Forscher vom Steinbeis-Institut in Heilbronn, wo man die Gründung und die Arbeit mehrerer Datengenossenschaften begleitet hat²¹, stützen die Erfahrungen in der Praxis und die Ergebnisse aus der Forschung die Annahme, dass eine institutionalisierte juristische Person ein Enabler für ein KMU-Ökosystem für den Datenaustausch sein kann. Neben der Bündelung von Ressourcen für Datenmanagement und Datenanalyse sowie der Möglichkeit, Skalen- und Verbundvorteile zu nutzen, könne eine solche Einrichtung sowohl als vertrauenswürdiger und neutraler Mittelsmann im Ökosystem und als Anker für eine ökosystemweite Datenverwaltung fungieren.²²

Die institutionalisierte Einheit ist eine eigene rechtliche Einheit, die zusätzliche administrative und unterstützende Aufgaben wie die Bereitstellung von Kapital übernehmen kann. Dazu gehört eine Kernrolle, die untrennbar damit verbunden ist: der Koordinator des Ökosystems. Der Koordinator ist verantwortlich für die IT-Ressourcen (entweder durch direkte Bereitstellung oder durch Beschaffung) und für die strategische Ausrichtung der Datenanalyseaktivitäten.

Als Rechtsform sei die deutsche Genossenschaft besonders geeignet. Allerdings erkennen die Autoren die Notwendigkeit an, andere Rechtsformen und Rahmenbedingungen zu prüfen, um zu einer Verallgemeinerung der Ergebnisse zu erlangen.

Als wesentlichen Vorteil einer Genossenschaft betrachten die Autoren das Element des Vertrauens. Der Datenaustausch in einer Genossenschaft

erfolgt nicht über anonyme Schnittstellen, die einem breiten Markt offen stehen, sondern innerhalb eines definierten Ökosystems von Kooperationspartnern, von denen keiner die Rolle eines zentralen Akteurs oder eines dominierenden Plattformanbieters übernehmen darf. Einige alternative Ideen zur Förderung der gemeinsamen Nutzung von Daten könnten den Ansatz jedoch ergänzen. So könnten beispielsweise neue Lösungen für die Bereitstellung von Datenräumen (z. B. das EU-Projekt GaiaX) einer Genossenschaft eine interessante technologische Grundlage und die Möglichkeit bieten, die Verbindung mehrerer Datengenossenschaften zu erleichtern.

Ausblick

Zum jetzigen Zeitpunkt ist schwer abzuschätzen, welche der hier vorgestellten Lösungsansätze den Bedürfnissen der KMU am besten entsprechen werden. Es hängt, wie so oft, von der jeweiligen Situation, dem spezifischen Anwendungsfall ab.

Mit Blick auf die Entwicklungen im Bereich des Confidential Computing, des föderierten Lernens und der Synthetischen Daten ist es durchaus vorstellbar, dass KMU von großen Datenpools profitieren können, ohne ihre Daten komplett aus der Hand zu geben. Jedoch sollte der Wartungsaufwand ebenso wie der Rechenaufwand nicht unterschätzt werden. Datenräume können ebenfalls ein sinnvoller Ansatz sein, sofern es gelingt, die Eintrittshürden, d. h. vor allem die Investitions- und Unterhaltungskosten, für KMU so gering wie möglich zu halten. Spezielle Marktplätze für (vortrainierte) KI-Modelle bieten für KMU eine weitere Möglichkeit, an die für ihre Zwecke geeigneten Algorithmen und Daten zu gelangen, wie es bei DIONE-X erprobt wird.

Auf der Organisationsebene ist das Modell der Datengenossenschaft für KMU nicht ohne Reiz. Der Aufwand für den Aufbau und den Geschäftsbetrieb der Genossenschaften sollte indes nicht unterschätzt werden.

Es ist gut möglich, dass in Zukunft weniger große Mengen an Maschinen-, Prozess-, Qualitäts- und Energiedaten, als vielmehr die daraus gewonnenen, aggregierten Erkenntnisse, wie in Gestalt von KI-Modellen, Algorithmen und verschlüsselten Datensätzen, ausgetauscht werden.

Für KMU ist das erst einmal eine gute Nachricht.

Weitere Informationen/Quellenangaben

- ¹ [Nutzung und Monetarisierung einer Industriellen Datenbasis](#)
- ² [How Smart, Connected Products Are Transforming Competition](#)
- ³ [Was ist ein Datenraum? Definition des Konzeptes Datenraum](#)
- ⁴ [Was ist ein Datenraum? Definition des Konzeptes Datenraum](#)
- ⁵ [Der Eclipse Dataspace Connector \(EDC\) – Architektur und Nutzen des Frameworks](#)
- ⁶ [Onboarding-Webinar für Interessierte und Neumitglieder](#)
- ⁷ [Industrial Edge Cloud](#)
- ⁸ [Zehn Partner gründen Joint Ventures Cofinity-X](#)
- ⁹ [Was ist Manufacturing X und was soll es bewirken?](#)
- ¹⁰ [Vorstudie Datenraum Manufacturing-X](#)
- ¹¹ [Vorstudie Datenraum Manufacturing-X](#)
- ¹² [Factory-X: Autonomie und Energieeffizienz in der vernetzten Produktion](#)
- ¹³ Jedoch gibt es laut der Berkeley-Professorin Raluca Ada Popa einige Unterschiede zwischen den Produkten der Anbieter, was die Geschwindigkeit und Integrität sowie die Benutzerfreundlichkeit angeht. Laut Popa hat die Intel SGX tendenziell stärkere Integritätsgarantien, während die AMD SEV-Enklave tendenziell schneller ist. Popa fügte hinzu, dass die Nitro-Enklaven von AWS größtenteils auf Software basieren und nicht das gleiche Maß an Hardwareschutz bieten wie Intel SGX. Intel SGX erfordert ein Code-Refactoring, um Legacy-Software auszuführen, während AMD SEV und Amazon Nitro-Enklaven besser für Legacy-Anwendungen seien, in: [Confidential computing provides revolutionary data encryption, UC Berkeley professor says](#)
- ¹⁴ [Wikipedia: Secure multi-party computation](#)
- ¹⁵ [Homomorphe Verschlüsselung und Europas Cloud: ein Baustein für Europas digitale Souveränität](#)
- ¹⁶ [Was sind die größten Herausforderungen und Grenzen der homomorphen Verschlüsselung in der Praxis?](#)
- ¹⁷ [Projekt „FLInK“ gestartet: Beschleunigter Transfer vortrainierter KI-Modelle in der fertigen Industrie](#)
- ¹⁸ [Is Synthetic Data the Future of AI?](#)
- ¹⁹ [Synthetic Data: A must have companion for data and AI led business growth and innovation](#)
- ²⁰ [Was sind Cloud-Plattformen für die Industrie?](#)
- ²¹ [Datengenossenschaft.com](#)
- ²² [Institutionalizing Analytic Data Sharing in SME Ecosystems – A Role-Based Perspective](#)

Gefördert durch:



Mittelstand-
Digital

aufgrund eines Beschlusses
des Deutschen Bundestages



Mittelstand-Digital
Zentrum
WertNetzWerke

Was ist Mittelstand-Digital?
Das Mittelstand-Digital Netzwerk bietet mit den Mittelstand-Digital Zentren und der Initiative IT-Sicherheit in der Wirtschaft umfassende Unterstützung bei der Digitalisierung. Kleine und mittlere Unternehmen profitieren von konkreten Praxisbeispielen und passgenauen, anbieterneutralen Angeboten zur Qualifikation und IT-Sicherheit. Das Bundesministerium für Wirtschaft und Klimaschutz ermöglicht die kostenfreie Nutzung der Angebote von Mittelstand-Digital. Weitere Informationen finden Sie unter www.mittelstand-digital.de.

Gefördert durch:



Mittelstand-
Digital 

aufgrund eines Beschlusses
des Deutschen Bundestages

So erreichen Sie uns!

Mittelstand-Digital Zentrum WertNetzWerke

Geschäftsstelle
c/o GS1 Germany GmbH
Maarweg 133
50825 Köln

www.mittelstand-digital-wertnetzwerke.de
+49 221 94714-439

Impressum:

Herausgeber: Mittelstand-Digital Zentrum WertNetzWerke

Autor: Ralf Keuper

Redaktion: Ulrich Hardt

Gestaltung/Produktion: Abeler Bollmann Werbeagentur GmbH, Wuppertal

www.mittelstand-digital-wertnetzwerke.de