

April 2019 · Dr. Stefan Heumann und Dr. Nicola Jentsch

Wettbewerb um Daten

Über Datenpools zu Innovationen



Think Tank für die Gesellschaft im technologischen Wandel

Executive Summary

Daten bilden die Grundlage für innovative Geschäftsmodelle und damit für die zukünftige Wettbewerbsfähigkeit von Unternehmen. Der Zugang der großen Internet-Plattformen zu den Daten ihrer Millionen Nutzer:innen wird im Wettbewerb als strategischer Vorteil von Unternehmen in den USA und China gesehen. Daten sind außerdem von äußerster Wichtigkeit für die Entwicklung von Künstlicher Intelligenz (KI), auch dieser Aspekt befeuert die kontroverse Diskussion um Datenzugang.

Die Politik hat erkannt, dass es dringend Antworten auf die Frage nach einem europäischen Modell einer Datenökonomie braucht. Die Einrichtung von Datenpools für die gemeinsame Nutzung und Verwertung von Daten spielt in dieser Debatte eine wichtige Rolle. Trotz fehlender, branchenübergreifender Plattformen und einer Vielzahl kleinerer Initiativen bei Forschungseinrichtungen und in der Privatwirtschaft, könnte sich aus der Kooperation der Akteure eine Stärke entwickeln. Dennoch bestehen viele offene Fragen, was den Aufbau von Datenpools angeht, sowie deren breitere Nutzung.

Mit diesem Papier wollen wir der Debatte über Datenkooperationen neue Impulse verleihen. Zu diesem Zweck haben wir mit einem Mapping bestehender Initiativen in Deutschland und wichtiger Initiativen in Europa begonnen. Darunter befinden sich Daten-Pools, Plattformen, Repositorien und andere Arten von Modellen aus der Privatwirtschaft und in geringerem Ausmaß auch der Forschung. Diese wurden nach verschiedenen Merkmalen klassifiziert. Das Mapping erhebt keinen Anspruch auf Vollständigkeit – es ist aber der erste Überblick über die vielfältige Landschaft der Initiativen, insbesondere in der Privatwirtschaft. Diese Übersicht zeigt auch, dass sich mittlerweile ein Ökosystem gebildet hat, bei dem es derzeit noch keine dominanten Player gibt.

Aus unserer Sicht ist ein breites Maßnahmenbündel notwendig, um Datenpools zum Durchbruch zu verhelfen. So wäre es unbedingt notwendig, Erkenntnisse über Datenpools aus der Forschung besser in die Umsetzung zu transferieren. Auch sollten die Stärken und Schwächen bestehender Modelle in der Praxis analysiert werden. Hierzu sollten Praktiker:innen und Forscher:innen in einen stärkeren Austausch treten und ihr Wissen teilen. Auf dieser Basis könnte ein entsprechendes Maßnahmenbündel entwickelt werden, um das Potenzial der Nutzung von Datenpools besser ausschöpfen zu können.

Eine erste Analyse zeigt, dass Unternehmen bei Datenkooperationen zögern, weil ihnen die Vorteile nicht ausreichend klar sind und die Risiken als schwer beherrschbar gesehen werden. Gerade in Bezug auf das Kartell- und Datenschutzrecht werfen Datenpools schwierige Fragen auf, die noch gelöst werden müssen, um für die notwendige Rechtssicherheit zu sorgen. Denn Datenkooperationen bei Preisen und Lieferantenbeziehungen könnten den Wettbewerb unterminieren. Hinzu kommt, dass Regeln für Datenpools entwickelt werden müssen, die Zugriffs- und Nutzungsrechte klären und überprüfbar machen und dabei geltendes Datenschutzrecht befolgen. Datenkooperationen können auch stark von neuen technischen Entwicklungen, insbesondere dezentralen Trainingsansätzen in der KI und der Produktion von synthetischen Daten, profitieren.

Insgesamt zeigt unsere Analyse, dass es dringend mehr Forschung über die benötigten Anreize und Rahmenbedingungen bedarf, um die Bildung von Datenpools von staatlicher Seite wirksam zu unterstützen. Voraussetzung für eine wirksame Datenpool-Strategie wäre zudem eine kritische Analyse existierender Datenpools. Bisher generiert nach unserem Kenntnisstand keine dieser Plattformen oder Pools das notwendige Wachstum, um den Datenmangel in Europa merklich abzufedern.

Danksagung

Wir bedanken uns für die exzellente Recherche von Anika Lange, unserer studentischen Hilfskraft. Außerdem möchten wir uns für den Austausch mit verschiedenen Personen bedanken, darunter Barbara Ebert (Rat für Informationsinfrastrukturen), Mathias Bug und Thomas Runge (beide RatSWD). Von den Verbänden Bitkom und BDI haben wir darüber hinaus Informationen zu bestehenden Datenpools erhalten, auch hierfür wollen wir uns bedanken.



Inhalt

I Einleitung	5
II Bunt, heterogen und jung: Sharing-Initiativen in Deutschland	7
2.1 Schnelle Typisierung von Sharing-Initiativen	8
2.2 Die Landschaft der Sharing-Initiativen in Deutschland	9
2.3 Anreize zur Teilnahme am Data Sharing	13
III Rechtliche Unsicherheiten im Datenaustausch	15
3.1 Wettbewerbsrecht	15
3.2 Datenschutzgrundverordnung (DSGVO)	18
IV Neue Ansätze für Daten-Initiativen	19
4.1 Data Governance	19
4.2 Dezentrale Verfahren	20
4.3 Synthetische Daten	21
V Schluss	22
Referenzen	24

I Einleitung

Im Wettbewerb um Innovationen durch Künstliche Intelligenz (KI) spielt der Zugang zu Daten eine zentrale Rolle. Derzeit dominieren die großen Tech-Konzerne aus dem Silicon Valley digitale Dienste und sammeln mit ihren Online-Plattformen viele wertvolle Daten. Die US-Konzerne nutzen die gesammelten Daten für die Entwicklung innovativer KI-Anwendungen, bei denen vor allem die Personalisierung ihrer Dienste und die Prognose von Präferenzen und Kaufinteressen ihrer Nutzer:innen im Mittelpunkt stehen. Sie drängen darüber hinaus zunehmend in Branchen vor, in denen die deutsche Industrie traditionell ihre Stärken hat, beispielsweise die Automobilbranche.

Europa steht bei der KI nicht nur im Wettbewerb mit den führenden Tech-Unternehmen aus den USA. Auch China hat im vergangenen Jahrzehnt erfolgreich eine eigene Digitalwirtschaft aufgebaut. Ursprünglich haben chinesische Unternehmen vor allem Geschäftsmodelle kopiert und lokal aufgesetzt. Mittlerweile haben sich daraus innovative digitale Ökosysteme entwickelt: Handelsplattformen, Bezahlssysteme, soziale Netze und Messenger-Dienste. Aufgrund des sehr großen Binnenmarktes verzeichnen diese Plattformen mittlerweile hohe Nutzer:innen-Zahlen. Die zentrale politische Frage für Europa lautet, wie zwischen den auf große Internetplattformen setzende Ansätzen aus China und USA ein „dritter Weg“ Europas gelingen kann, der digitale Innovationen und Wertschöpfung aus Daten ermöglicht, aber gleichzeitig unseren Werten und Interessen folgt und die Grund- und Menschenrechte schützt.

Die deutsche Regierung hat den Wettbewerb angenommen. Die Strategie Künstliche Intelligenz der Bundesregierung setzt stark auf die Verbesserung des Transfers von Forschung in kommerzielle Anwendungen. Und auch die jüngst von Wirtschaftsminister Peter Altmaier vorgestellte nationale Industriestrategie 2030 steht ganz im Zeichen der Sorge um die zukünftigen Wettbewerbsfähigkeit der deutschen Wirtschaft angesichts der Innovationskraft der großen Online-Plattformen bei digitalen Geschäftsmodellen sowie Entwicklung und Einsatz von KI.¹ Allerdings bleibt die Konkretisierung des europäischen Ansatzes im globalen Wettbewerb um Innovation und neue Technologien wie KI eine große Herausforderung.

Im Zentrum dieser Herausforderung stehen Daten. Eine Industriepolitik, die auf digitale Geschäftsmodelle und KI setzt, muss sich mit der Verfügbarkeit

¹ Bundesministerium für Wirtschaft und Energie (BMWi), Nationale Industriestrategie 2030, https://www.bmwi.de/Redaktion/DE/Publikationen/Industrie/nationale-industriestrategie-2030.pdf?__blob=publicationFile&v=10

und Nutzbarmachung von Daten befassen. Hierbei gilt es, den bestehenden Rechtsrahmen, darunter Datenschutz-, Datenbankwerk- und Urheberrechte, einzuhalten und gleichzeitig die Entwicklung datengetriebener Geschäftsmodelle und mit ihnen verknüpfter Technologien zu fördern.

In Berlin war Datenpolitik lange kein Thema, mittlerweile ist sie eine zentrale industriepolitische Herausforderung geworden. Open Data – vor kurzer Zeit noch ein Spezialthema für Netzaktivist:innen – findet sich in den KI-Strategien Deutschlands und Frankreichs ebenso wieder wie im KI-Strategieprozess auf EU-Ebene. Da viele, auch gesellschaftlich wertvolle Datensätze in Privatbesitz sind, ist mittlerweile eine Diskussion darüber entbrannt, ob diese durch ein Daten-für-alle-Gesetz² allgemein zugänglich gemacht werden sollten.³ In der Debatte geht es vor allem darum, wie und ob eine nationale Dateninfrastruktur für die Privatwirtschaft und für staatliche Leistungen aufgebaut werden soll.

In der Diskussion geht oft unter, dass Unternehmen bereits heute Datenkooperationen eingehen können. Viele Unternehmen suchen derzeit nach neuen Zugangsmöglichkeiten zu Daten oder arbeiten an kooperativen Datennutzungsmodellen. Mit solchen Modellen lassen sich Lieferketten effizienter organisieren oder vernetzte Forschungsk Kooperationen aufsetzen. Datenpools werden für die Entwicklung einer wettbewerbsfähigen Datenökonomie in Deutschland ein wichtiger Treiber sein. Sie sind daher zunehmend auch von industriepolitischer Bedeutung. In der Wirtschaft und der Forschung werden die Potentiale des Datenteilens erkannt, aber es fehlt derzeit an einer übergeordneten wirtschaftspolitischen Strategie, wie Datenpools auf- und ausgebaut werden können.

Um eine solche Strategie zu entwickeln, bedarf es im ersten Schritt einer kritischen Bestandsaufnahme. Mit diesem Impuls wollen wir hierzu einen Beitrag leisten. Diese erste Übersicht zeigt eine überraschende Anzahl und Vielfalt an Datenpools, Repositorien und Plattformen in Deutschland. Diese haben wir anhand verschiedener Merkmale klassifiziert, beispielsweise ob

2 SPD-Parteivorstand (2019). Digitaler Fortschritt durch ein Daten-für-Alle-Gesetz, <https://www.spd.de/aktuelles/daten-fuer-alle-gesetz/>

3 Kritik am SPD-Vorschlag kommt unter anderem von Nadine Schön, Vize-Chefin der CDU/CSU-Bundestagsfraktion: Handelsblatt, Union äußert Vorbehalte gegen ‚Daten-für-alle‘-Konzept von Nahles, <https://www.handelsblatt.com/politik/deutschland/digitalisierung-union-aeussert-vorbehalte-gegen-daten-fuer-alle-konzept-von-nahles/23984464.html?ticket=ST-884082-hZ9RgERHyHkTIqCS67zx-ap4> Grundsätzliche Kritik hat auch Stefan Heumann im Tagesspiegel geäußert. Online verfügbar hier: <https://www.stiftung-nv.de/de/publikation/daten-fuer-alle-innovation-fuer-wenige>

es sich um einen horizontalen Austausch unter Wettbewerbern handelt oder um einen vertikalen zwischen Zulieferern und Abnehmern.

Mit diesem Impuls wollen wir Erkenntnisse aus der Recherche teilen und erste Empfehlungen für die Politik abgeben. Ausgangslage ist, dass wir statt einer Konzentration bei wenigen großen Plattformen eine fragmentierte Datenlandschaft über viele Unternehmen hinweg beobachten. Daher ist die Kooperation bei der Nutzung von Daten so wichtig. Unsere Recherche zeigt, dass es hierzu auch bereits eine Vielzahl von Ansätzen gibt. Keinem dieser Ansätze ist bisher der große Durchbruch gelungen. Derzeit generiert keine der Plattformen oder Pools ein exponentielles Wachstum, soweit wir dies anhand der uns vorliegenden Informationen einschätzen können. Über eine smarte Kombination von ökonomischen Anreizen könnte die Teilnahme an Datenpools deutlich erhöht werden. Hierzu muss allerdings auch die Rechtssicherheit im Datenschutz und Kartellrecht verbessert und entsprechend eine Governance für Teilen und Austausch von Daten entwickelt werden.

II Bunt, heterogen und jung: Sharing-Initiativen in Deutschland

Bislang fehlte es in Deutschland an einer Analyse des Status quo. Das bedeutet aber nicht, dass in Deutschland noch keine Datenpools und Repositorien existieren. Die Landschaft der Sharing-Initiativen in Deutschland ist vielfältig und unübersichtlich. Das Problem des „Kartographierens“ einer solch heterogenen Landschaft wurzelt im Fehlen von Begriffsdefinitionen. Derzeit gibt es keine anerkannten Definitionen verschiedener Sharing-Typen, darunter Plattformen, Pools oder Repositorien. Jede Art des *Landscaping* leidet gezwungenermaßen unter dieser Unschärfe. Die hier vorgestellten Erkenntnisse sind aus diesem Grund auch nur als erster Überblick ohne Anspruch auf Vollständigkeit zu verstehen.

Wir definieren Datenpools als Einrichtungen, über die verschiedene Partner (in der Mehrzahl Unternehmen) ihre Daten zentral oder dezentral verwalten und deren Ziel es ist, diese Daten gemeinsam zu nutzen und zu verwerten. „Repositorien“ nehmen wir ebenfalls auf, da diese gleichfalls eine gemeinschaftliche Datenverwaltung darstellen können. Dieser Begriff wird synonym für Pools verwendet. Derzeit gibt es keine allgemeingültigen Definitionen dieser Begriffe.

Insgesamt haben wir rund 60 Initiativen identifiziert, die verschiedene Varianten des *Data Sharings* darstellen. Sie reichen von Forschungs-Pools wie

dem SoWiDataNet zu Industrie-Plattformen wie den International Data Spaces (ehemals Industrial Data Space). Der Schwerpunkt hier sind privatwirtschaftliche Plattformen mit einer Ausrichtung auf unternehmenszentriertes *Sharing* von Daten. Forschungsinitiativen beziehen wir nur „am Rande“ zum Zwecke des Überblicks ein. Die bereits geleistete Arbeit anderer Stakeholder in diesem Bereich sollte nicht repliziert werden.⁴

Auch in der Privatwirtschaft gibt es erste Übersichten. So haben der Bitkom-Verband und der Bundesverband der Deutschen Industrie Listen erstellt.⁵ Diese Listen konnten wir mit unserer Datenbank abgleichen.

Vereinzelt existieren bereits andere Untersuchungen zu dem Thema. Beispielsweise befragte PricewaterhouseCoopers (2018) deutschlandweit rund 200 Firmen bezüglich ihres Datenaustauschs. 57 Prozent gaben an, dass sie regulär Daten weitergeben. Der größere Teil der Firmen tut dies mit Kunden (83 Prozent) und etwa 15 Prozent der befragten Firmen auch mit Konkurrenten.⁶ Zu den Datentypen, die geteilt werden, gehören Lieferstatus- und Lagerdaten oder Source Codes von Programmen. In dieser Umfrage zeigt sich auch, dass große Unternehmen häufiger als KMUs Daten teilen, dies aber zumeist mit Kunden oder Zulieferern tun.

2.1 Schnelle Typisierung von Sharing-Initiativen

Zunächst lassen sich die Modelle in öffentlich oder privat finanzierte Initiativen unterscheiden. Die wichtigsten Merkmale der *Sharing*-Initiativen sind in Tabelle 1 aufgelistet. Unter den Merkmalen der Initiativen findet sich zum einen die Finanzierung (öffentlich/privat/Public Private Partnership (PPP)). Zum anderen haben wir geprüft, ob es sich bei einer Initiative um einen „horizontalen Austausch“ zwischen Stakeholdern auf derselben Produktionsstufe (z.B. zwischen zwei Autoherstellern) handelt oder um einen „vertikalen Austausch“ auf unterschiedlichen Produktionsstufen (z.B. zwischen Zulieferer und Abnehmer).

4 Rat für Informationsinfrastrukturen (2016). Leistung aus Vielfalt – Empfehlungen zu Strukturen, Prozessen und Finanzierung des Forschungsdatenmanagement in Deutschland, <http://www.rfii.de/download/rfii-empfehlungen-2016/>

5 Die Autor:innen bedanken sich bei den Verbänden für die Weitergabe der Listen.

6 So sind Forschungsverbände auch im Wettbewerbsgesetz gestattet.



Tabelle 1. *Sharing*-Initiativen: Merkmale

Merkmale	Merkmalsausprägungen			
Finanzierung	Öffentlich	Privat	PPP	
Austausch	Horizontal	Vertikal	Beides	
Orientierung	Forschung	Privatwirtschaftlich		
Funktion	B2B	B2C	P2P	C2B
Art	Datenmarktplatz	Datenpool / Sammlung		
Nutzung	intern	extern		
Zweck	Verschiedene Nennungen			
Status	Aufbau	laufend	beendet	
Organisation	Zentrale Datenspeicherung		Dezentrale Datenspeicherung	
Reichweite	Deutschland	Europa	Global	
Analysetools	Bereitstellung von verschiedenen Tools zu Analyse der Daten (Nennungen)			
Anreize	Verschiedene Nennungen			

Anmerkung: PPP - Public Private Partnership, B2B – Business to business, B2C – Business to consumer, P2P – Person to person und C2B – Consumer to business.

Manche der Initiativen qualifizieren wir als Datenpool oder Daten-Sammlung. Diese können zentraler (Daten werden an einem Ort zusammengeführt) oder dezentraler Natur sein (Daten verbleiben bei den Unternehmen und werden nicht zusammengeführt). Als Datenmarktplätze bezeichnen wir alle Einrichtungen, die ein Monetarisierungsmodell bezüglich des Zugriffs auf Daten integriert haben. Dies ermöglicht Unternehmen, über diese Einrichtungen Einkommensströme aus ihren Daten-Assets zu generieren.

Manche Initiativen werden vornehmlich für interne Zwecke genutzt, wie die Optimierung von Lieferprozessen, andere für externe Kooperationen (s. Zeile ‚Nutzung‘). Gerade das vertikale Sharing bietet oft einen ersten Einstieg für Unternehmen ins Teilen von Daten: Die Kostenvorteile der Verbesserung von Lieferprozessen liegen auf der Hand.

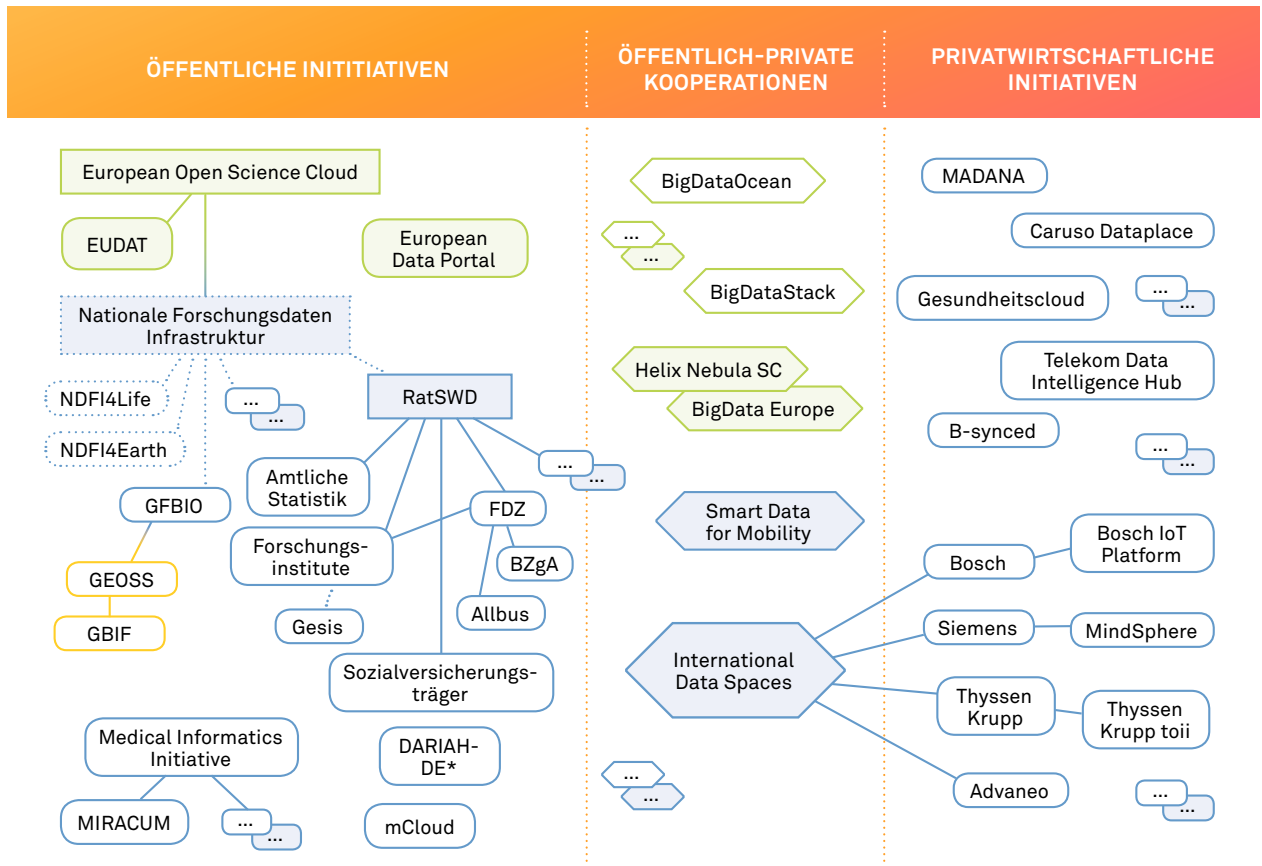
2.2 Die Landschaft der Sharing-Initiativen in Deutschland

Schaubild 1 ist eine aktuelle Mind Map der Sharing-Initiativen in Deutschland, inklusive wichtiger EU-Initiativen und inklusive der geplanten Initiativen (wie die Nationale Forschungsdateninfrastruktur). Auf der linken Seite sind die öffentlichen Plattformen, mittig die öffentlich-privaten Kooperati-



onen und rechts die privatwirtschaftlichen Initiativen positioniert. Nur die wichtigsten Initiativen sind hier verzeichnet, zur Übersichtlichkeit verzichten wir auf die Darstellung aller der von uns identifizierten Modelle. Informationsgrundlage für diese Recherche sind die Webseiten der Initiativen gewesen und in geringerem Ausmaß Telefoninterviews zur Klärung offener Fragen.

Schaubild 1 Landschaft der Sharing-Initiativen



Legende: Grün = europäische Initiativen, orange = globale Initiativen, blau = deutsche Initiativen bzw. mit Sitz in Deutschland. Formen mit Füllung markieren sehr große Initiativen mit branchenübergreifendem Infrastrukturcharakter.

Die verschiedenen Initiativen sind teils untereinander verbunden, aber über Kooperationen oder gemeinsame Mitglieder und derzeit nicht über einen Datenaustausch. Ein paar größere Initiativen haben mit ihrem übergreifenden Infrastrukturcharakter das Potential von ‚Gravitationszentren‘ (z.B. den International Data Spaces). Die Verbindungslinien markieren Teilnehmer oder

Kooperationspartner. So sind Bosch und Siemens Kooperationspartner der International Data Spaces, betreiben aber auch eigene Plattformen.

Andere Projekte sind Infrastrukturen von Infrastrukturen: Sie verbinden mehrere Daten-Initiativen miteinander. Dies gilt vor allem für die initiierte Nationale Forschungsdateninfrastruktur (NFDI).⁷

Andere bieten Standards und Protokolle.⁸ So hat der RatSWD das Modell des ‚Forschungsdatenzentrums‘ etabliert. Es handelt sich dabei unter anderem um eine Abstufung des Datenzugriffs basierend auf der Sensibilität der Daten. Diese Protokolle sind von über 30 Forschungsdatenzentren übernommen worden.⁹

Die Mind Map zeigt auch diejenigen Bereiche, in denen derzeit keine Verbindungen existieren. So ist derzeit keine Daten-Kooperation der großen Industriepattformen mit Forschungsdatenzentren erkennbar.

In Deutschland gibt es derzeit Industriekooperationen wie *Smart Data for Mobility* (SD4M) für die Verkehrswirtschaft und branchenübergreifende den International Data Spaces. Die SD4M-Plattform hat zum Ziel, Daten aus unterschiedlichen Quellen (Unternehmen, öffentliche Daten) zusammenzuführen, durch semantische Interoperabilität zu strukturieren und perspektivisch auch zu monetarisieren.

Die International Data Spaces sollen Unternehmen den sicheren Austausch von Daten ermöglichen. Ursprünglich als Industrial Data Space gegründet, war die anfängliche Entwicklung dieser Datenaustauschplattform eng mit der Industrie 4.0-Plattform verknüpft. Die Weiterentwicklung zum International Data Spaces soll eine breitere Akteurslandschaft über die Industrie und die Landesgrenzen hinaus ansprechen. Innerhalb der Initiative werden Standards und Protokolle zu Datenaustausch und Nutzung erprobt.

Andere Initiativen befördern Datenaustausch durch die Standardisierung von Metadaten oder stellen Suchfunktionen für Datensätze Dritter bereit. In

⁷ Hier bereiten verschiedene Zentren derzeit Konsortiansanträge vor.

⁸ In der Fachliteratur werden auch gemeinsame Standards und Protokolle als Plattform gesehen.

⁹ Siehe Webseite des RatSWD <https://www.ratswd.de/forschungsdaten/akkreditierung>

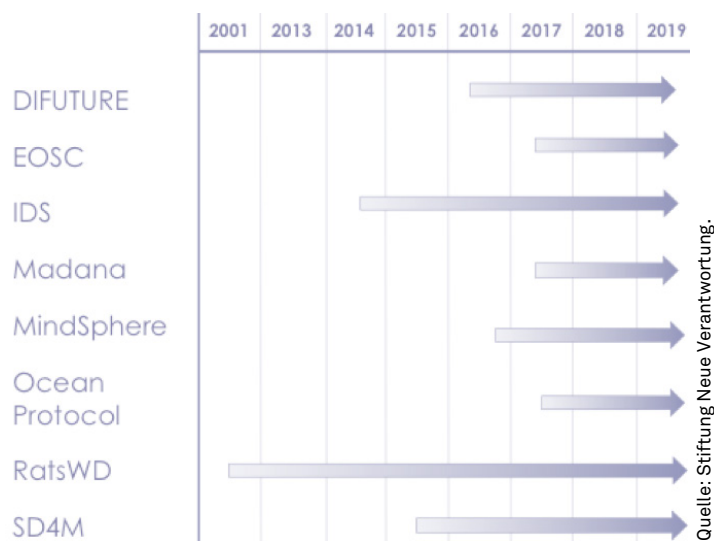


dieses Segment gehört beispielsweise re3data.org. Diese Initiativen sind für die Standardisierung sehr wichtig.

Rund 64 Prozent der betrachteten Plattformen und Sharing-Initiativen sind öffentlich finanziert, etwa 34 Prozent sind rein privatfinanziert und nur 2 Prozent sind als PPP aufgesetzt. Die Mehrheit der Initiativen (80 Prozent) sind als Datenpools und Sammlungen ausgestaltet, etwa 20 Prozent als Datenmarktplätze. Letztere implementieren einen Mechanismus zu Monetarisierung von Datensätzen oder von Analysen aus den Daten.

Unter den verschiedenen Datenmarktplätzen, Beispiele sind Madana oder Caruso Dataplace, hat sich bislang noch kein dominanter Marktplatz entwickelt. Es gibt also derzeit kein „Amazon für Daten“. Die Startups verfolgen unterschiedliche Modelle: Manche verbinden Blockchain-Technologien mit Verschlüsselungs- und Depersonalisierungs-Verfahren. In einer vertrauenswürdigen Infrastruktur soll verhindert werden, dass über die Daten Rückschlüsse auf interne Unternehmensprozesse gezogen werden können. Von den insgesamt 13 Datenmarktplätzen konnten wir sieben auf Basis ihrer Website-Informationen als dezentral einordnen. Beispielsweise schreibt die Telekom über ihren Data Intelligence Hub: „Dabei bleibt die Telekom stets neutraler Treuhänder (...) und gewährleistet die jeweilige Datensouveränität durch eine dezentrale Datenhaltung.“ Aus Platzgründen gehen wir hier nicht weiter auf Open Data Initiativen ein, obwohl diese auch als Datenpools betrachtet werden könnten. So stellt z.B. die Govdata-Plattform ein interessantes Beispiel eines Pools für Daten aus der öffentlichen Hand dar.¹⁰

Schaubild 2 Existenz von Initiativen und Plattformen



¹⁰ Webseite: <https://www.govdata.de/>

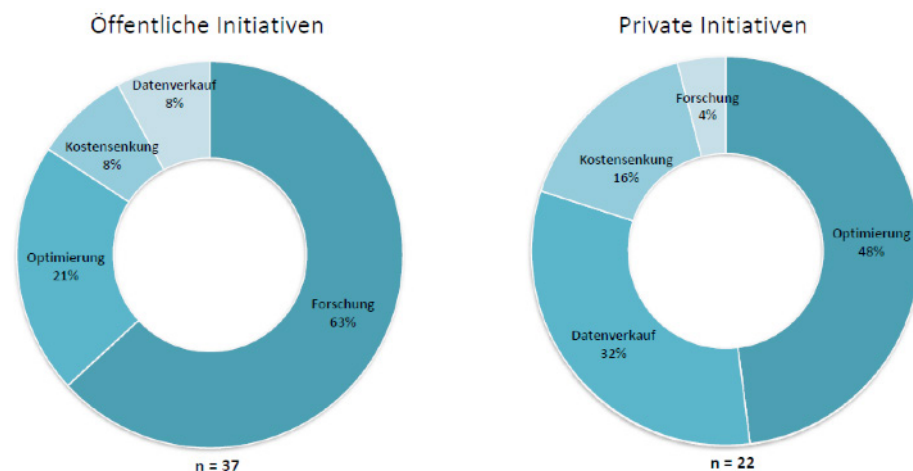


Schaubild 2 zeigt für ausgewählte Plattformen, seit wann diese existieren. Kaum eine der von uns identifizierten Initiativen ist älter als zehn Jahre. Gerade im Forschungsbereich sind viele der Initiativen kaum nachhaltig. Die Daten werden für ein bestimmtes Projekt erstellt und nach Beendigung archiviert oder vernichtet. Schon vielfach wurde hier Anreizsetzung sowie eine Verstetigung der Projekte gefordert.¹¹

2.3 Anreize zur Teilnahme am Data Sharing

Es existieren unterschiedliche Motive für das Teilen von Daten. Diese spiegeln sich in den Zielsetzungen der analysierten Dateninitiativen wider. Während bei den öffentlich finanzierten Pools zumeist Forschungszwecke im Vordergrund stehen (s. Schaubild 3), sind Optimierung von Prozessen und Produkten oder Datenverkauf die wichtigsten Gründe für die Teilnahme an privaten Initiativen.

Schaubild 3 Teilnahme-Anreize: Öffentliche und Private Initiativen



Anmerkung: Für nicht alle Initiativen waren Informationen verfügbar. Quelle: Stiftung Neue Verantwortung

Zwar sind die Möglichkeiten der datengetriebenen Innovation für Unternehmen sehr attraktiv, ermöglicht die Teilnahme an einem Pool doch die Erweiterung des eigenen Datenportfolios. Viele Unternehmen sehen in der Bereitstellung ihrer Daten aber berechtigterweise auch ein Sicherheitsrisiko. Zum einen könnten Daten Unternehmensgeheimnisse reflektieren, zum anderen könnten sie Hacker:innen einladen, neue Wege der Infiltrierung zu finden.

¹¹ Rat für Informationsinfrastrukturen (2016). Leistung aus Vielfalt – Empfehlungen zu Strukturen, Prozessen und Finanzierung des Forschungsdatenmanagement in Deutschland, <http://www.rfii.de/download/rfii-empfehlungen-2016/>

Mit einer steigenden Anzahl an externen Datenbanken steigt außerdem das Risiko der Reidentifikation aus depersonalisierten Daten.

Vertikale Austauschprozesse, die internen Zwecken dienen, sind meist der Beginn zu mehr Offenheit und Kooperation von Unternehmen mit externen Partnern. Auch vertikalen Austausch haben wir deshalb in unsere Betrachtung einbezogen. Bei vertikalen Austauschprozessen sind die Kosten- und Effizienzvorteile am deutlichsten ersichtlich (s. auch PricewaterhouseCoopers 2018), weshalb diese Art des Datenaustauschs, also mit Kunden und/oder Zulieferern, häufiger vorkommt.

Welche Wettbewerbseffekte ergeben sich aus der Beteiligung an einem Datenaustausch? Die überraschende Antwort ist, dass sich Unternehmen *nicht immer* besser stellen. Je nach Fall kommt die ökonomische Theorie zu sehr unterschiedlichen Resultaten. Diese hängen von verschiedenen Marktbedingungen ab. Ursachen können die Asymmetrie in den Marktanteilen der Unternehmen,¹² die geteilten Datentypen¹³ oder die Privatsphären-Präferenzen der Datensubjekte sein.¹⁴

Beispielsweise könnten bei asymmetrischen Marktanteilen kleine Unternehmen relativ mehr von einer Daten-Plattform profitieren als große. Dies war in der Vergangenheit beispielsweise in manchen Schwellenländern in der Branche der Kreditauskunfteien für große Banken das Argument, ihre Daten nicht mit kleineren Finanzinstituten zu teilen.

Ein Teilen von Kundendaten kann außerdem einen Wettbewerbseffekt haben, wenn der Wettbewerb um einzelne Kunden:innen dadurch sehr viel härter wird. In diesem Falle sinken die Gewinne der Unternehmen. Sollten Unternehmen von trägen Kunden, die nicht schnell zwischen den Anbietern wechseln, höhere Preise verlangen können, dann können sie sinkende Gewinne allerdings kompensieren.¹⁵ Der Nettoeffekt ist letztendlich eine empirische Frage und macht die Gewinnauswirkungen des Teilens von Daten zu einer

12 Liu, Q. und Serfes, K. (2006). Customer Information Sharing among rival firms, *European Economic Review* 50: 1571 – 1600.

13 Jentzsch, N., Sapi, G., Suleymanova, I. (2013). Targeted Pricing and Customer Data Sharing Among Rivals, *International Journal of Industrial Organization* 31, 131-144; Jappelli T., M. Pagano (1993). Information sharing in credit markets, *Journal of Finance* 48 (5): 1693-1718

14 Akcura, M.T. und K. Srinivasan (2005). Research Note: Customer Intimacy and Cross-Selling Strategy, *Management Science* 51 (4): 1007 – 1012.

15 Diese Erkenntnisse gelten für den Fall, dass Kund:innendaten über die Flexibilität der Kund:innen im Anbieterwechsel geteilt werden.

unternehmensintern sehr schwierigen strategischen Frage, die quasi nur fallabhängig beantwortet werden kann.

Wettbewerbseffekte entstehen auch dann, wenn es sich um sogenannte ‚*massive blind networks*‘ handelt, also Netze, in denen die Teilnehmer die Identität der anderen Teilnehmer nicht sehen können. Dies schützt zwar die Datenquellen, *trotzdem* können die bereitgestellten Daten zu Markteintritten oder mehr Wettbewerb in unterschiedlichen Kundensegmenten führen. Da es sich um eine in ihren Auswirkung schwer prognostizierbare aber strategisch sehr wichtige Entscheidung für Unternehmen handelt, zögern diese verständlicherweise mit der Teilnahme an solchen Initiativen. Die angesprochenen wettbewerblichen Anreizprobleme existieren außerdem in Kombination mit rechtlichen Unsicherheiten, die dem Datenschutz- und Wettbewerbsrecht entspringen.

Insbesondere kleine und mittlere Unternehmen haben Probleme damit zu entscheiden, unter welchen Bedingungen der Datenaustausch für das Unternehmen von Vorteil ist. Es könnte daher angezeigt sein, Beratungsangebote zu schaffen.

III Rechtliche Unsicherheiten im Datenaustausch

3.1 Wettbewerbsrecht

Im Kartellrecht wird eine Informationsweitergabe unter konkurrierenden Unternehmen mit dem Zweck, gemeinsam die Marktunsicherheit zu senken, mit empfindlichen Strafen geahndet. Dies bedeutet, dass ein Informationsaustausch ein Verstoß gegen Art. 101 AEU-Vertrag (ehemals Art. 81 EG-Vertrag) darstellen kann, wenn er die Marktunsicherheit reduziert oder beseitigt und dadurch „tatsächlich oder möglicherweise zu einer spürbaren Einschränkung des Wettbewerbs zwischen den Unternehmen“ führt. Die Beurteilung, ob Datenaustausch eine solche Auswirkung hat, ist nicht trivial. In der Vergangenheit haben Behörden, um einen Verstoß festzustellen, eine Reihe von Faktoren zur Entscheidungsgrundlage genommen (Tabelle 2).¹⁶

¹⁶ Obst, S. (2008). Informationsaustausch in Verbänden aus Sicht der Europäischen Kommission, Kartellrechtliches Symposium, Bonn, 30. Oktober 2008.

Außerdem gibt es Hinweise des Bundeskartellamts bezüglich der Abwägung der wettbewerbsfördernden oder beschränkenden Wirkung.¹⁷

Tabelle 2. Informationsaustausch zwischen Konkurrenten

Kriterien	Problematisch	Unproblematisch
Aggregationsebene	Individuell	Aggregiert
Charakter	Geschäftsgeheimnisse	Öffentliche Information
Informationsalter	Prognosen	Historisch
Frequenz des Austausches	Häufig	Selten
Konzentrationsgrad	Oligopol	Zersplittert
Homogenität der Produkte	Hoch	Niedrig
Symmetrie Marktanteile	Symmetrisch	Asymmetrisch
Existierende Transparenz	Niedrig	Hoch

Quelle: Obst (2008) und OECD (2010).

Es gab aber bereits Ausnahmen aus dem generellen Kartellierungsverbot.¹⁸ Und mittlerweile existieren auch gerichtliche Klarstellungen zum Informationsaustausch. So hat der Europäische Gerichtshof beispielsweise für die Finanzbranche festgestellt, dass Finanzinstitute individuelle Schuldnerinformationen austauschen können, da dieser Austausch nicht dem Zweck dient, die Marktunsicherheit zu senken.¹⁹ Vielmehr werden durch den Austausch solcher Informationen systemische Finanzrisiken reduziert.

¹⁷ Bundeskartellamt (2017). Big Data und Wettbewerb – Schriftenreihe „Wettbewerb und Verbraucherschutz in der digitalen Wirtschaft“, Oktober, http://www.bundeskartellamt.de/SharedDocs/Publikation/DE/Schriftenreihe_Digitales/Schriftenreihe_Digitales_1.pdf?__blob=publicationFile&v=3

¹⁸ Im April 2017 ist beispielsweise die Gruppenfreistellungsverordnung für die Versicherungswirtschaft ausgelaufen.

¹⁹ European Court of Justice (2006). Case C-238/05 Asnef-Equifax, Servicios de Información sobre Solvencia y Crédito, SL and Administración del Estado v Asociación de Usuarios de Servicios Bancarios (Ausbanc), European Court Reports 2006 I-11125, <http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A62005CJ0238>.

Mittlerweile hat die Kommission Leitlinien zur horizontalen Zusammenarbeit (Horizontal-GVO) und zur vertikalen Zusammenarbeit (Vertikal-GVO) veröffentlicht.²⁰ In der Horizontal-GVO wird festgehalten, dass Ausnahmeregelungen bestehen können. Dafür müssen aber vier Voraussetzungen erfüllt werden:

- die Vereinbarung führt zu wirtschaftlichen Vorteilen, sie führt zu einer Verbesserung der/des Produktion/Vertriebs oder zur Förderung des technischen oder wirtschaftlichen Fortschritts (z.B. in Form von Effizienzgewinnen);
- die Vereinbarung ist notwendig, um diese Effizienzgewinne zu erreichen;
- Verbraucher:innen müssen in angemessener Weise an den Effizienzgewinnen beteiligt werden;
- die Vereinbarung eröffnet den Parteien nicht die Möglichkeit, für einen wesentlichen Teil der betroffenen Produkte den Wettbewerb auszuschalten.

Es wird in den Leitlinien auch festgehalten, dass es unterschiedliche Formen des Informationsaustauschs gibt, darunter den direkten unter Wettbewerbern und den indirekten über Dritte, also zum Beispiel eine Plattform. Sobald Unternehmen aber etwas über die Wettbewerbsstrategie der Konkurrenten lernen können, haben Daten „Kollusionspotential.“ In den Fokus der Aufsicht kommen hier vor allem Daten über geplantes künftiges Preis- oder Mengenverhalten unter Wettbewerbern. Der Austausch anderer Arten von Informationen wird anhand der tatsächlichen Auswirkungen auf den Wettbewerb bewertet.²¹

Hierbei ist anzumerken, dass Forschungsk Kooperationen auch unter dem derzeit existierenden Wettbewerbsgesetz möglich sind.

Zusammengefasst bedeutet dies, dass sich Unternehmen nicht sicher sein können, dass der Austausch anderer Daten (als Mengen und Preise) *immer unproblematisch* und damit legal ist. Hier gibt es – rechtlich gesehen – keinen sicheren Hafen.

20 Horizontal-GVO (<https://eur-lex.europa.eu/legal-content/DE/TXT/?uri=LEGISSUM%3A126062>) und Vertikal-GVO (<https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2010:102:0001:0007:DE:PDF>).

21 Horizontal-GVO (<https://eur-lex.europa.eu/legal-content/DE/TXT/?uri=LEGISSUM%3A126062>).

3.2 Datenschutzgrundverordnung (DSGVO)

Der Anwendungsbereich der Datenschutzgrundverordnung, die seit Mai 2018 in Kraft ist, bezieht sich auf personenbeziehbare und pseudonymisierte Daten. Daten, die mit Individuen nicht in Zusammenhang gebracht werden können, fallen nicht unter den Anwendungsbereich der Verordnung. „Personenbeziehbarkeit“ ist genau genommen ein Qualitätsgrad. Dies bedeutet, dass es unterschiedliche Grade von Personenbeziehbarkeit gibt. Aus diesem Grund bestehen trotz umfangreicher Erklärungen durch die Art. 29-Working Party²² erhebliche Unsicherheiten, ab wann Daten rechtssicher depersonalisiert sind. Die Frage, wie ‚individualisiert‘ Daten sind, spielt auch im Wettbewerbsrecht eine Rolle (s. Tabelle 2, Stichwort „Aggregationsebene“).

Unter Anonymisierung wird ein irreversibler Prozess verstanden, bei dem kein Personenbezug mehr hergestellt werden kann. Bei der Pseudonymisierung der Daten können die Daten über einen Schlüssel auf das Datensubjekt zurückbezogen werden. Eine relativ robuste Depersonalisierung kann quasi nur über eine Kombination von mehreren Verfahren erreicht werden, weil jedes Einzelverfahren für sich genommen Stärken und Schwächen hat.²³

Mit diesen Rechtsunsicherheiten geht einher, dass es keine automatisierten Verfahren gibt, die es erlauben, personenbeziehbare Daten in Datenbanken zu erkennen und ‚herauszufiltern‘. Es kann daher auch nicht automatisch erkannt werden, welche Daten potentiell geteilt werden könnten und welche nicht.²⁴ Diese Unsicherheiten existieren also neben den eben diskutierten unternehmensstrategischen Unsicherheiten.

Es gibt eine Reihe neuer Ansätze, wie diesen Herausforderungen in Zukunft begegnet werden kann, darunter synthetische Daten und dezentrale Systeme. Auf diese wollen wir in aller Kürze eingehen.

²² Art 29-Working Party (2014). Opinion 05/2014 on Anonymisation Techniques adopted on 10 April 2014. https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf

²³ Art 29-Working Party (2014). Opinion 05/2014 on Anonymisation Techniques adopted on 10 April 2014. https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf

²⁴ Es gibt eine Reihe von Software-Tools, die Compliance in Unternehmen mit der DSGVO sicherstellen, dabei handelt es sich um die sogenannten *Privacy Management Systems*. Künftig müsste recherchiert werden, inwiefern diese Software solche Prozesse automatisieren kann.

IV Neue Ansätze für Daten-Initiativen

Zunächst muss in Unternehmen oder Datenpools eine Data Governance aufgebaut werden. Es handelt sich dabei um eine Unternehmenspolitik, die den Umgang mit Daten regelt. Innerhalb einer solchen Governance können Unternehmen mit Datenanalysen und Kollaborationsansätzen experimentieren.

Eine Governance und neue Verfahren, wie beispielsweise das dezentrale Lernen, können das *Sharing* attraktiver machen. Auf der technologischen Ebene werden die Systeme so aufgesetzt, dass sie Barrieren des Datenaustausches abbauen, dies wird weiter unten beispielsweise anhand der synthetischen Daten erklärt.

4.1 Data Governance

Die verschiedenen Daten-Initiativen, die wir untersucht haben, unterliegen jeweils einer eigenen Data Governance. Data Governance ist ein Begriff aus dem Unternehmens-Management. Er bezeichnet Richtlinien, Protokolle, Strukturen und Handlungsanweisungen, auf deren Basis Unternehmen oder Daten-Initiativen Entscheidungen bezüglich Daten und Datenanalysen fällen. Insbesondere legt sie fest, wer Entscheidungsautorität besitzt, welche Nutzergruppen existieren, wie Zugriffsrechte verteilt werden und welche Sicherheitsmaßnahmen implementiert werden. Innerhalb von Unternehmen soll die Data Governance verhindern, dass unkontrollierte Datensätze und Datenanalysen entstehen. Bei personenbeziehbaren Daten oder pseudonymisierten Daten sollte der Datenaustausch grundsätzlich *nicht ohne* Data Governance stattfinden, da Dokumentation eine wichtige Pflicht der Datenkontrolleur:innen ist und unkontrollierten Entwicklungen vorgebeugt werden muss.

Data Governance muss allerdings nicht nur auf der Unternehmensebene, sondern auch übergeordnet auf der Ebene der Datenpools entwickelt werden. Im Kern handelt es sich dabei um ein Regelwerk, das festlegt, wie Daten in den Pool eingespeist bzw. Teilnehmern zur Verfügung gestellt werden können und wer, unter welchen Bedingungen, auf welche Daten zugreifen und was mit den Daten gemacht werden darf. Die Data Governance muss, um wirksam zu sein, entsprechend technisch implementiert. Hierbei handelt es sich vor allem um die technische Umsetzung der in der Governance festge-

legten Zugangs- und Nutzungsregeln. So braucht es zum Beispiel technische und semantische Standards für das Einspeisen und Teilen der Daten.

Hierbei existieren auch schon Praktiken. Beispielsweise existieren Datenzugangsprotokolle, die von über 30 Forschungsdatenzentren übernommen wurden.

Auch Metastandards sind wichtig. Hier haben wir in unserer Recherche eine ganze Reihe von Standard-Initiativen gefunden, darunter DublinCore oder MARC21. Zusätzlich ist eine entsprechende technische Dokumentation der Datenzugriffe notwendig, um die Einhaltung der Regeln überprüfen.

Im Rahmen dieser ersten Übersicht haben wir die Data-Governance-Ansätze der Datenpools nicht systematisch erfasst. Hier besteht mehr Bedarf für eine Bestandsaufnahme und Analyse, denn gute Data-Governance-Modelle sind zwingend erforderlich, um der Entwicklung und Skalierung von Datenpools zum Durchbruch zu verhelfen.

4.2 Dezentrale Verfahren

Es gibt eine ganze Reihe von Verfahren, mit deren Hilfe innovative Data Governance Ansätze entwickelt und implementiert werden können. So haben Forscher:innen bereits vor Jahren gezeigt, dass es möglich ist, Prognosen von Neuronalen Netzen gemeinsam zu verbessern, ohne dass die Inputdaten geteilt werden müssen. Es handelt sich dabei um das sogenannte Kollaborative Lernen (*collaborative deep learning*) oder Verteiltes Lernen (*distributed learning*). Experimente zeigen, dass die Präzision der gemeinsam gelernten Modelle über die der individuell gelernten Modelle steigt. Das Verfahren funktioniert ungefähr wie folgt:

1. Jeder Teilnehmer trainiert sein Modell (Netz) auf seinem lokalen Datensatz;
2. Während des Trainings werden ausgewählte Parameter des Modells oder deren Aktualisierungen mit den anderen Teilnehmern geteilt;²⁵
3. Gemeinsam wird die Präzision der einzelnen Modelle gesteigert.

Weder die *Ergebnisse* (z.B. *Klassifikationsentscheidung*) der Modelle, noch die *Inputdaten* werden mit den anderen geteilt. Dies ist das eigentlich Attraktive an dieser Lösung: Modelle können verbessert werden, aber die Da-

²⁵ Shokri, R. und V. Shmatikov (2015). Privacy-preserving Deep Learning, CCS'15, October 12–16, 2015, Denver, Colorado, USA <https://www.comp.nus.edu.sg/~reza/files/Shokri-CCS2015.pdf>

ten verbleiben lokal beim ‚Besitzer.‘ Zusätzlich können die ausgetauschten Daten leicht verrauscht werden.²⁶

Solche Verfahren können nicht nur auf verschiedenen Datentypen (Audio, Text, Video) angewandt werden, sondern auch für verschiedene Typen Neuroner Netze. Die ‚Privatheit‘ von Millionen von Modellparametern in Tiefen Neuronalen Netzen und die gemeinsame Entwicklung analytischer Werkzeuge, ist eine Forschungsrichtung, die sich zunehmender Popularität erfreut.²⁷

4.3 Synthetische Daten

Andere Verfahren setzen direkt bei der Datenproduktion an. In verschiedenen Institutionen existieren sehr gute Erfahrungen mit der Produktion und dem Einsatz synthetischer Daten.²⁸ Synthetische Daten sind eine Art „Repräsentation“ der Originaldaten. Sie werden aber nicht durch direkte Messung am Datensubjekt erhoben, sondern durch mathematische Verfahren aus den Originaldaten gewonnen.²⁹ Diese Daten gelten als depersonalisiert, sofern mit ihnen verbundene Experimente der Reidentifikation entsprechend negativ ausgefallen sind.

Der Vorteil gegenüber traditionellen Depersonalisierungsmethoden – wobei Synthetisierung mit diesen kombiniert werden kann – besteht darin, dass die Datenqualität erhalten bleibt. Andere Verfahren reduzieren durch das Einstreuen von Zufallsfaktoren die Datenqualität und können sogar Beziehungen, die zwischen Variablen bestehen, zerstören. Dies ist nicht der Fall bei der Synthetisierung.

26 Geyer, R.C., T. Klein und M. Nabi (2018). Differentially Private Federated Learning: A Client Level Perspective, <https://arxiv.org/pdf/1712.07557.pdf>

27 In Deutschland gibt es unter anderem eine Privacy Preserving Machine Learning am Max-Planck-Institut für Intelligente Systeme in Tübingen (<https://privacy-preserving-machine-learning.github.io/>)

28 Ein Beispiel ist in den USA das U.S. Census Bureau und in Deutschland das Institut für Arbeitsmarktforschung.

29 Hierzu ausführlich: Drechsler, J. und N. Jentzsch (2018). „Synthetische Daten: Innovationspotential und gesellschaftliche Herausforderungen“ Stiftung Neue Verantwortung (SNV), Mai, https://www.stiftung-nv.de/sites/default/files/synthetische_daten.pdf

V Schluss

Für die deutsche Wirtschaft sind Datenkooperationen von großer strategischer Bedeutung. Durch sie kann die Großindustrie, aber auch der Mittelstand, in die Lage versetzt werden, Wettbewerbsnachteile gegenüber den großen Datengiganten aus den USA und China zumindest teilweise wettzumachen. Hierzu gibt es in Deutschland und Europa erste Ansätze. Ziel unseres Impulses ist es, mehr Aufmerksamkeit in Wirtschaft und Politik auf diese Initiativen zu lenken. Unsere Analyse macht auch deutlich, dass eine Vielzahl von Maßnahmen notwendig sind, um das Potenzial von Datenkooperationen besser heben zu können.

So ist dringend mehr Forschung über bestehende Datenpools notwendig, um Stärken und Schwächen bestehender Initiativen zu identifizieren und vertiefend zu analysieren. Hierzu sollten Praktiker:innen und Forscher:innen in einen stärkeren Austausch treten und ihr Wissen teilen. Auf Basis einer solchen Bestandsaufnahme und dem Austausch zentraler Stakeholder kann dann ein entsprechendes Maßnahmenbündel entwickelt werden, um das Potenzial der Nutzung von Datenpools ausschöpfen zu können. Die Konturen dieses Maßnahmenbündels lassen sich bereits auf Basis unserer ersten Recherchen erkennen.

Die Ursachen für die positive Entwicklung, aber auch für das Scheitern von Datenkooperationsmodellen müssen genauer untersucht werden. Dabei sollte ein besonderer Schwerpunkt auf die Analyse der ökonomischen Anreizstrukturen und der bestehenden Hürden für die Teilnahme an solchen Kooperationen gelegt werden. Zusätzlich ist mehr Forschung zur Data Governance von Daten-Pools notwendig. Das Ziel sollte hierbei sein, Best Practices und Standards sowohl auf technischer wie auch auf rechtlicher und organisatorischer Ebene zu identifizieren und weiterzuentwickeln. Datenkooperationen können auch stark von neuen technischen Entwicklungen, insbesondere dezentralen KI-Ansätzen und synthetischen Daten, profitieren.

Zusätzlich benötigen Datenkooperationen auch einen klaren Rechtsrahmen, der allen Beteiligten ein möglichst hohes Maß an Rechtssicherheit bringt. Hierzu müssen vor allem Haftungs- und Datenschutzfragen mit Bezug zum Kartellrecht analysiert werden.

Die Analysen und Forschungsergebnisse bieten auch eine gute Grundlage für einen stärkeren Dialog zwischen relevanten Stakeholdern (Unternehmen, Datenplattformen, Datenschützer:innen, Forschung, Politik, etc.) zu Chancen



Dr. Stefan Heumann und Dr. Nicola Jentzsch

April 2019

Wettbewerb um Daten

und Risiken von Datenkooperationen. Dieser Dialog ist dringend notwendig, um die Potenziale von Datenkooperationen erkennen und mögliche Hürden ausräumen zu können. Hierin besteht auch eine große Chance, den europäischen Weg bei Digitalisierung und KI zwischen China und USA zu konkretisieren und mit Leben zu füllen. Wir hoffen, dieser Debatte mit diesem Papier einen Impuls geben zu können.

Referenzen

Akcura, M.T. und K. Srinivasan (2005). Research Note: Customer Intimacy and Cross-Selling Strategy, *Management Science* 51 (4): 1007 – 1012.

Art 29-Working Party (2014). Opinion 05/2014 on Anonymisation Techniques adopted on 10 April 2014. https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf

Bundeskartellamt (2017). Big Data und Wettbewerb – Schriftenreihe „Wettbewerb und Verbraucherschutz in der digitalen Wirtschaft“, Oktober, http://www.bundeskartellamt.de/SharedDocs/Publikation/DE/Schriftenreihe_Digitales/Schriftenreihe_Digitales_1.pdf?__blob=publicationFile&v=3

Bundesministerium für Wirtschaft und Energie (BMWi), Nationale Industriestrategie 2030, https://www.bmwi.de/Redaktion/DE/Publikationen/Industrie/nationale-industriestrategie-2030.pdf?__blob=publicationFile&v=10

Drechsler, J. und N. Jentzsch (2018). “Synthetische Daten: Innovationspotential und gesellschaftliche Herausforderungen” Stiftung Neue Verantwortung (SNV), Mai, https://www.stiftung-nv.de/sites/default/files/synthetische_daten.pdf

European Court of Justice (2006). Case C-238/05 Asnef-Equifax, Servicios de Información sobre Solvencia y Crédito, SL and Administración del Estado v Asociación de Usuarios de Servicios Bancarios (Ausbanc), *European Court Reports* 2006 I-11125, <http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A62005CJ0238>.

Geyer, R.C., T. Klein und M. Nabi (2018). Differentially Private Federated Learning: A Client Level Perspective, <https://arxiv.org/pdf/1712.07557.pdf>

Neuerer, D. (2018). Union äußert Vorbehalte gegen ‚Daten-für-alle‘-Konzept von Nahles, <https://www.handelsblatt.com/politik/deutschland/digitalisierung-union-aeussert-vorbehalte-gegen-daten-fuer-alle-konzept-von-nahles/23984464.html?ticket=ST-884082-hZ9RgERHyHkTlqCS-67zx-ap4>

Horizontal-GVO (<https://eur-lex.europa.eu/legal-content/DE/TXT/?u->

[ri=LEGISSUM%3A126062](#)) und Vertikal-GVO (<https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2010:102:0001:0007:DE:PDF>).

Jentzsch, N., Sapi, G., Suleymanova, I. (2013). Targeted Pricing and Customer Data Sharing Among Rivals, *International Journal of Industrial Organization* 31, 131-144; Jappelli T., M. Pagano (1993). Information sharing in credit markets, *Journal of Finance* 48 (5): 1693-1718.

Liu, Q. und Serfes, K. (2006). Customer Information Sharing among rival firms, *European Economic Review* 50: 1571 – 1600.

Markl, V. (2019). Eine nationale Daten- und Analyseinfrastruktur als Grundlage digitaler Souveränität, *Informatik Spektrum*; Preprint https://www.bbdc.berlin/fileadmin/news/documents/Datenanalyseinfrastruktur_VM_20181119_pre.pdf

Obst, S. (2008). Informationsaustausch in Verbänden aus Sicht der Europäischen Kommission, *Kartellrechtliches Symposium, Bonn, 30. Oktober 2008*.

OECD (2010), *Information Exchanges Between Competitors under Competition Law*, OECD Policy Roundtables, <http://www.oecd.org/competition/cartels/48379006.pdf>

PricewaterhouseCoopers (2018) *Data exchange as a first step towards data economy*, Report, <https://www.pwc.de/en/digitale-transformation/data-exchange-as-a-first-step-towards-data-economy.pdf>

Rat für Informationsinfrastrukturen (2016). *Leistung aus Vielfalt – Empfehlungen zu Strukturen, Prozessen und Finanzierung des Forschungsdatenmanagement in Deutschland*, <http://www.rfii.de/download/rfii-empfehlungen-2016/>

Rat für Informationsinfrastrukturen (2016). *Leistung aus Vielfalt – Empfehlungen zu Strukturen, Prozessen und Finanzierung des Forschungsdatenmanagement in Deutschland*, <http://www.rfii.de/download/rfii-empfehlungen-2016/>

Shokri, R. und V. Shmatikov (2015). *Privacy-preserving Deep Learning*, CCS'15, October 12–16, 2015, Denver, Colorado, USA <https://www.comp.nus.edu.sg/~reza/files/Shokri-CCS2015.pdf>



Dr. Stefan Heumann und Dr. Nicola Jentsch

April 2019

Wettbewerb um Daten

SPD-Parteivorstand (2019). Digitaler Fortschritt durch ein Daten-für-Alle-Gesetz, <https://www.spd.de/aktuelles/daten-fuer-alle-gesetz/>



Über die Stiftung Neue Verantwortung

Think Tank für die Gesellschaft im technologischen Wandel

Neue Technologien verändern Gesellschaft. Dafür brauchen wir rechtzeitig politische Antworten. Die Stiftung Neue Verantwortung ist eine unabhängige Denkfabrik, in der konkrete Ideen für die aktuellen Herausforderungen des technologischen Wandels entstehen. Um Politik mit Vorschlägen zu unterstützen, führen unsere Expertinnen und Experten Wissen aus Wirtschaft, Wissenschaft, Verwaltung und Zivilgesellschaft zusammen und prüfen Ideen radikal.

Über die Autor:innen

Stefan Heumann ist Politikwissenschaftler und Mitglied des Vorstands. Er arbeitet, spricht und schreibt zu einer Vielzahl von Themen der nationalen und internationalen Digitalpolitik. Zurzeit liegt einer seiner Interessenschwerpunkte auf den wirtschaftlichen, gesellschaftlichen und politischen Fragestellungen rund um Künstliche Intelligenz. Neben Papieren bei der SNV hat er zahlreiche digitalpolitische Analysen und Gastbeiträge in renommierten nationalen und internationalen Medien und Zeitungen veröffentlicht.

Nicola Jentzsch war Projektleiterin für den Bereich Datenökonomie bei der Stiftung Neue Verantwortung. Im März 2019 schied sie auf eigenen Wunsch aus der SNV aus.

So erreichen Sie den Autor

Dr. Stefan Heumann
Mitglied des Vorstands
sheumann@stiftung-nv.de
+49 (0)30 81 45 03 78 80

Impressum

Stiftung Neue Verantwortung e. V.

Beisheim Center
Berliner Freiheit 2
10785 Berlin

T: +49 (0) 30 81 45 03 78 80

F: +49 (0) 30 81 45 03 78 97

www.stiftung-nv.de

info@stiftung-nv.de

Design:

Make Studio

www.make-studio.net

Layout:

Johanna Famulok

Free Download:

www.stiftung-nv.de



Dieser Beitrag unterliegt einer CreativeCommons-Lizenz (CC BY-SA). Die Vervielfältigung, Verbreitung und Veröffentlichung, Veränderung oder Übersetzung von Inhalten der Stiftung Neue Verantwortung, die mit der Lizenz „CC BY-SA“ gekennzeichnet sind, sowie die Erstellung daraus abgeleiteter Produkte sind unter den Bedingungen „Namensnennung“ und „Weiterverwendung unter gleicher Lizenz“ gestattet. Ausführliche Informationen zu den Lizenzbedingungen finden Sie hier:

<http://creativecommons.org/licenses/by-sa/4.0/>