

Arnold, René; Hildebrandt, Christian; Taş, Serpil

Research Report

Europäische Datenökonomie: Zwischen Wettbewerb und Regulierung. Endbericht

WIK-Consult Studie

Provided in Cooperation with:

WIK Wissenschaftliches Institut für Infrastruktur und Kommunikationsdienste GmbH, Bad Honnef

Suggested Citation: Arnold, René; Hildebrandt, Christian; Taş, Serpil (2020) : Europäische Datenökonomie: Zwischen Wettbewerb und Regulierung. Endbericht, WIK-Consult Studie, WIK-Consult GmbH, Bad Honnef

This Version is available at:

<https://hdl.handle.net/10419/251537>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Europäische Datenökonomie: Zwischen Wettbewerb und Regulierung

Endbericht

Autoren:

Dr. René Arnold
Christian Hildebrandt
Serpil Taş

WIK-Consult GmbH
Rhöndorfer Str. 68
53604 Bad Honnef

Bad Honnef, Januar 2020

Impressum

WIK-Consult GmbH
Rhöndorfer Str. 68
53604 Bad Honnef
Deutschland
Tel.: +49 2224 9225-0
Fax: +49 2224 9225-63
E-Mail: info@wik-consult.com
www.wik-consult.com

Vertretungs- und zeichnungsberechtigte Personen

Geschäftsführerin	Dr. Cara Schwarz-Schilling
Direktor Abteilungsleiter Post und Logistik	Alex Kalevi Dieke
Direktor Abteilungsleiter Netze und Kosten	Dr. Thomas Plückebaum
Direktor Abteilungsleiter Regulierung und Wettbewerb	Dr. Bernd Sörries
Leiter der Verwaltung	Karl-Hubert Strüver
Vorsitzende des Aufsichtsrates	Dr. Daniela Brönstrup
Handelsregister	Amtsgericht Siegburg, HRB 7043
Steuer-Nr.	222/5751/0926
Umsatzsteueridentifikations-Nr.	DE 123 383 795

Inhalt

Executive Summary	i
1 Einleitung	1
2 Daten richtig nutzen	4
2.1 Mehrschichtiges Referenzmodell	4
2.2 Definition des Begriffs "Daten"	7
2.3 Charakterisierung von Daten aus wirtschaftlicher Sicht	13
2.4 Datentypologien	15
2.5 Datenmenge versus -qualität?	19
3 Wertschöpfung aus Daten	26
3.1 Datenökonomie	26
3.2 Geschäftsmodelle in der Datenökonomie	30
3.3 Innovation in der Datenökonomie	35
4 Wettbewerb in der Datenökonomie	46
4.1 Rolle der Daten für den Wettbewerb	46
4.2 Datengetriebene Marktmacht und Markteintrittsbarrieren	50
4.3 Förderung des Wettbewerbs durch Transparenz und Zugang zu offenen Daten	57
5 Ansätze zur gemeinsamen Nutzung von Daten	61
5.1 Datenmärkte und Handel mit Daten	61
5.2 Horizontaler Datenaustausch und vertikales Datenpooling	64
5.3 Statische Datenportabilität und dynamische Interoperabilität	68
6 Implikationen für die Wirtschaftspolitik und Regulierung	74
6.1 Politisch relevante Herausforderungen der Datenökonomie	75
6.2 Referenzarchitekturen als Weg in die Zukunft	78
7 Fazit	85
Literatur	87

Links zu den wichtigsten Ergebnissen

Die wichtigsten Ergebnisse von Kapitel 2	3
Die wichtigsten Ergebnisse aus Kapitel 3	25
Die wichtigsten Ergebnisse aus Kapitel 4	45

Die wichtigsten Ergebnisse von Kapitel 5	60
Die wichtigsten Ergebnisse von Kapitel 6	73

Executive Summary

Daten und ihre ökonomischen Auswirkungen durchdringen alle Bereiche der Wirtschaft. Die sogenannte Datenökonomie ist kein neuer Wirtschaftszweig, sondern eher eine ständige Herausforderung für alle Unternehmen, im Rahmen einer neuen Dimension der Wertschöpfung wettbewerbsfähig und innovativ zu sein.

Die Ergebnisse der vorliegenden Studie widersprechen der üblichen Sichtweise, dass der Zugang zu Daten die größte Herausforderung für eine florierende Datenökonomie in Europa ist. Eine viel dringlichere Herausforderung ist das Fehlen einer gemeinsamen "Datensprache", die den Datenaustausch erleichtern kann. Eine gemeinsame Datensprache würde helfen, Daten von einem Kontext in einen anderen zu übertragen, um Datensätze zu erstellen, die von anderen Unternehmen genutzt werden können. Sogenannte "Referenzarchitekturen" bieten genau diese gemeinsame Sprache.

Die politischen Entscheidungsträger in ganz Europa sollten die Entwicklung und Einführung solcher Referenzarchitekturen fördern, um die Quantität und Qualität des Datenaustauschs durch Datenteilung und Datenpooling zwischen Unternehmen zu erhöhen.

Es hat sich gezeigt, dass die Verpflichtung zur Datenteilung nur von eingeschränktem Nutzen ist und sich sogar als nachteilig erweisen kann, insbesondere wenn keine entsprechenden Referenzarchitekturen vorhanden sind.

Förderung einheitlicher Datenarchitekturen zur Begünstigung der europäischen Datenökonomie

Im digitalen Bereich führende Länder wie Estland und Südkorea haben ihren Erfolg auf einheitlichen (zugrunde liegenden) Referenzarchitekturen aufgebaut, die den Datenaustausch und die Datenwiederverwendung erleichtern. Diese beiden Länder haben mehr als andere vom sozioökonomischen Potenzial der Daten profitiert.

Da Daten in allen Wirtschaftszweigen eine immer wichtigere Rolle spielen, legen die Ergebnisse dieser Studie europäischen Entscheidungsträgern nahe, die Entwicklung und Einführung einheitlicher Referenzarchitekturen zu fördern. Diese Architekturen stellen einen technologieneutralen und sektorübergreifenden Ansatz dar, der es sowohl kleinen als auch großen Unternehmen ermöglichen wird, im Wettbewerb zu bestehen und das wirtschaftliche Potenzial der Datenerhebung in einer zunehmend digitalisierten Welt innovativ zu nutzen.

Der Zugang zu Daten scheint für eine wachsende Datenökonomie kein größeres Hindernis zu sein, da die Möglichkeiten der Datenerfassung, -erhebung und -analyse deutlich verbessert wurden. Schwierig ist es für die Unternehmen jedoch, die vorhandenen Datensätze zu entdecken und ihre Eignung für die Erreichung ihrer eigenen wirtschaftlichen Ziele zu beurteilen. Referenzarchitekturen können diesen Prozess erleichtern, da sie einen Rahmen bieten, um potentielle Anbieter relevanter Datensätze ausfindig zu

machen und genügend zusätzliche Informationen (Metadaten) über die Datensätze enthalten, so dass Unternehmen beurteilen können, ob ein bestimmter Datensatz oder Teile davon dem jeweiligen Zweck entsprechen.

Wann immer ein bereits vorhandener Datensatz verwendet werden kann, ist der Zugriff auf diesen wahrscheinlich effizienter als die eigene Datenerfassung. In diesen Fällen sollten Unternehmen, die ihren Gewinn optimieren wollen, dem Datenaustausch gegenüber der Datenerfassung den Vorzug geben. Dafür spricht auch, dass Unternehmen sowohl in horizontalen Datenteilungs-Vereinbarungen als auch in vertikalen Datenpooling-Systemen wie dem Industrial Data Space bereits häufig Daten austauschen. Ausgehend von dieser Prämisse dürften Referenzarchitekturen, sobald sie verfügbar sind und branchenübergreifend eingesetzt werden, die Quantität und Qualität des Datenaustauschs für große und kleine Unternehmen in vielen Branchen erhöhen.

Wirtschaftsakteure, nicht die politischen Entscheidungsträger, sollten über den Nutzen eines Datenaustauschs entscheiden

Ob der Zugriff auf Daten Dritter überhaupt geeignet ist, eine bestimmte betriebliche Aufgabe zu bewältigen, sollten die beteiligten Wirtschaftsakteure nach eigenem Ermessen entscheiden. Wie unsere Studie verdeutlicht, sind Daten, die in einem bestimmten Kontext zu einem bestimmten Zweck erfasst wurden, möglicherweise nicht für einen anderen Kontext oder einen anderen Zweck geeignet. Folglich muss ein Unternehmen von Fall zu Fall entscheiden, ob eine eigene Datenerfassung, der Zugriff auf Daten Dritter oder ein gemischter Ansatz die beste Lösung ist. Dies hängt natürlich auch davon ab, ob ein anderes Unternehmen geeignete Daten erfasst und bereit ist, die Bedingungen für den Zugang Dritter zu diesen Daten auszuhandeln. Einheitliche Datenarchitekturen können die Kosten für die Aufbereitung der Daten für einen erfolgreichen Austausch senken, so dass für ein Unternehmen, das geeignete Daten erfasst, die Hürden für den Eintritt in entsprechende Verhandlungen abgebaut werden. Solche Datenarchitekturen können Lizenzbestimmungen integrieren, die sicherstellen, dass die Daten nach dem Austausch nicht über den vereinbarten Zweck hinaus verwendet werden. Sie können darüber hinaus Funktionen beinhalten, die die Auffindbarkeit potenzieller Datenanbieter verbessern.

Eine Verordnung über die gemeinsame Nutzung von Daten wird Bedenken in Bezug auf die Datenökonomie kaum ausräumen

Der branchenübergreifende Datenaustausch ist ein wichtiger Schritt hin zu einer florierenden europäischen Datenökonomie. In jüngster Zeit sind allerdings Bedenken hinsichtlich einer Datenkonzentration bei nur wenigen großen Unternehmen in die politische Debatte eingebracht worden. Diese Bedenken betreffen (1) die Bindung der Verbraucher an die Diensteanbieter (Lock-in-Effekt) und (2) insbesondere kleine und mittlere Unternehmen (KMU), die aufgrund fehlenden Datenzugangs nicht wettbewerbsfähig seien.

Um den **Lock-in-Effekt** zu verringern, wurde im Rahmen der Allgemeinen Datenschutzverordnung (General Data Protection Regulation (GDPR)) die Datenportabilität festgeschrieben. Obwohl der Prozess der Datenextraktion überschaubar ist, stellt der Transfer dieser extrahierten (personenbezogenen) Daten zu einem anderen Dienstleister nach wie vor eine große Herausforderung für die Datenökonomie insgesamt dar. Darüber hinaus können regulatorische Vorgaben zur Dateninteroperabilität eine erfolgreiche Datenportierung erleichtern. Private Initiativen wie das Data Transfer Project sind entstanden, um dieser Herausforderung zu begegnen.

Die Autoren dieser Studie sind der Ansicht, dass der **Datenzugang** weder der einzige noch der entscheidendste Einflussfaktor für den Geschäftserfolg ist. Grundsätzlich ist ein unzureichender Datenzugang nur schwer vorstellbar. Praktisch jeder Datenpunkt (*Datum*) kann mit verschiedenen Methoden erfasst oder anderweitig bestimmt werden, so dass ein Unternehmen, das nicht in der Lage ist, Daten mit dem gleichen Verfahren wie sein Wettbewerber zu erfassen, die erforderlichen Daten möglicherweise noch auf eine andere Art und Weise erfassen kann.

Vor diesem Hintergrund trifft die Annahme, dass (personenbezogene) Daten eine wesentliche Einrichtung sind im gleichen Sinne wie (physische) Infrastruktur (z.B. die Eisenbahn oder Telekommunikationsnetze), nicht zu. Um als "wesentliche" Einrichtung oder Infrastruktur zu gelten, müssten (personenbezogene) Daten zwei Bedingungen erfüllen: 1) der Marktzutritt zum komplementären Markt ist ohne Zugang zu dieser Einrichtung praktisch nicht möglich; und 2) ein Anbieter auf dem komplementären Markt kann diese Einrichtung nicht mit vertretbarem Aufwand reproduzieren und es gibt kein Substitut. Es ist offensichtlich, dass (personenbezogene) Daten keine der beiden Bedingungen erfüllen. Die erste Bedingung setzt voraus, dass ein Unternehmen im Vergleich zu anderen eine höhere Effizienz der Kostenstrukturen aufweist. Jedes Unternehmen der Datenökonomie kann jedoch beliebige (personenbezogene) Daten zu oft vernachlässigbaren Kosten erheben. Folglich gibt es bei (personenbezogenen) Daten kein Monopol, und somit ist die erste Bedingung für eine wesentliche Einrichtung nicht erfüllt. Was die zweite Bedingung betrifft, so können (personenbezogene) Daten immer mit vertretbarem Aufwand dupliziert oder reproduziert werden, auch wenn dafür die Zustimmung des Endnutzers erforderlich ist. Daher können (personenbezogene) Daten keine wesentliche Einrichtung sein.

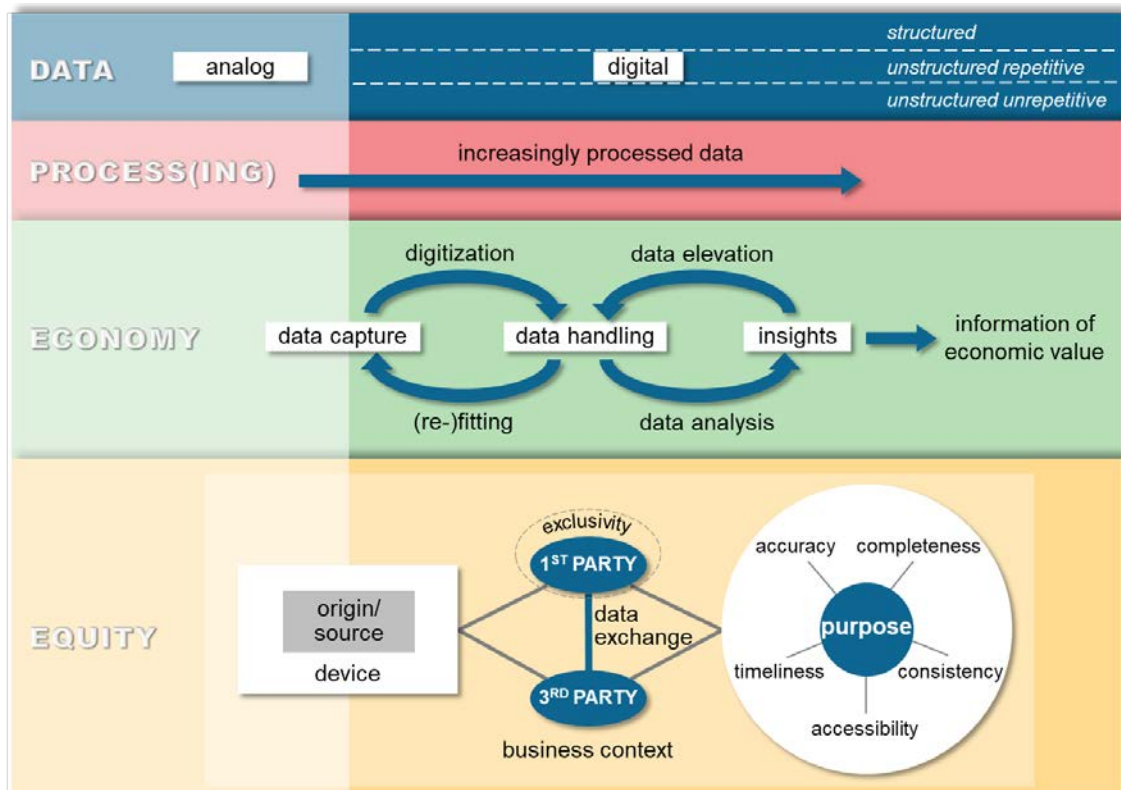
Eine allgemeine gesetzliche Verpflichtung zur gemeinsamen Nutzung oder zum Pooling von Daten ist daher nur schwer zu rechtfertigen. In der Praxis ist die Generierung von wirtschaftlichem Wert aus Daten (Big Data) ein recht komplexer und vielschichtiger Prozess. Selbst wenn also eine regulatorische Verpflichtung gerechtfertigt wäre, müsste diese ausreichend flexibel sein, um den unterschiedlichen Gegebenheiten gerecht zu werden. Allerdings wäre damit wahrscheinlich ein hoher Verwaltungsaufwand verbunden, der auf die Unternehmen abgewälzt würde, und könnte daher der europäischen Datenökonomie letztlich mehr schaden als nützen.

Ohne eine einheitliche Datenarchitektur dürfte die Belastung durch Transaktionskosten jeglichen Anreiz zur Nachfrage nach Datenaustausch zunichtemachen. Sollten personenbezogene Daten ausgetauscht werden, müsste das Konzept der Einwilligung gemäß der Datenschutzgrundverordnung geändert werden, da ein solcher Austausch fast zwangsläufig eine Änderung des Zwecks, zu dem die Daten ursprünglich erfasst wurden, mit sich bringen würde. Es ist schwer vorstellbar, wie die zuständigen Behörden eine solche regulatorische Verpflichtung ohne vollständige Transparenz der von allen beteiligten Parteien gespeicherten Daten, ihrer Eigenschaften, Kontexte und Zwecke umsetzen könnten. Schließlich würde ein solcher Markteingriff unweigerlich Gewinner und Verlierer hervorbringen - ohne Nachweis eines Marktversagens oder eines wirtschaftlichen Schadens.

Falsche Annahmen tragen zu einem verzerrten Narrativ über Daten und ihre sozioökonomischen Auswirkungen bei

Die aktuelle Debatte über die sozioökonomischen Auswirkungen von Daten und die Regulierung der gemeinsamen Datennutzung unterliegt teilweise falscher Annahmen über die Eigenschaften der Daten und der Funktionsweise der Datenökonomie. Diese Studie soll die Debatte mit entsprechender Fachkenntnis aufwerten, indem sie die wichtigsten Merkmale von Daten und die Rolle beleuchtet, die sie für Unternehmen, Innovation und Wettbewerb spielen. Hierfür wurde ein mehrschichtiges Modell entworfen, mit dem die Bedeutung von Daten erklärt werden soll (s. Abbildung 2–1).

Referenzmodell zur Darstellung der (wirtschaftlichen) Eigenschaften von Daten



Quelle: WIK-Consult.

Das Modell verdeutlicht wesentliche Eigenschaften von Daten und zeigt verschiedene Unzulänglichkeiten und Fehleinschätzungen der gängigen Narrative zur Datenökonomie auf. Die erste der vier Schichten stellt die Umwandlung von **analogen Daten** (Signale aus der realen Welt) in **digitale Daten** dar. Auf dieser Datenschicht können digitale Daten weiter in **strukturierte, unstrukturierte, sich wiederholende** und **unstrukturierte, sich nicht wiederholende** digitale Daten unterteilt werden. Während andere Arten der Darstellung je nach Kontext nützlich sein können, geben sie doch durchweg keinen Aufschluss über die Datenbeschaffenheit und ihre Auswirkungen auf die Datennutzung. Dies ist aber besonders wichtig, da die überwiegende Mehrheit der heute verfügbaren digitalen Daten unstrukturiert ist (z.B. Video, Bilder, Text, natürliche Sprache, Social Media Streams).

Die **Prozessschicht** macht deutlich, dass es keine digitalen Rohdaten gibt. Bereits bei der Digitalisierung wurde darüber entschieden, welche analogen Signale in digitale Bits umgewandelt werden sollen und wie genau diese Umwandlung erfolgt (z.B. welche Sensoren oder Schnittstellen eingesetzt werden). Der Verarbeitungsgrad steigt in der Regel mit der weiteren Nutzung der digitalen Daten an. Diesem Umstand steht die weit verbreitete Vorstellung gegenüber, dass digitale Daten objektive Fakten sind, die daher für jeden Marktteilnehmer den gleichen Wert haben.

Die **Wirtschaftsschicht** zeigt, dass die Umwandlung von Daten in **wirtschaftlich verwertbare Informationen** nicht unbedingt ein einfacher Prozess ist. Erst in Kombination mit einem bestimmten Kontext können Daten zu wirtschaftlich verwertbaren Informationen werden. Logischerweise erfolgt die Digitalisierung während der **Datenerfassung**, wobei Geräte mit Sensoren und/oder Schnittstellen eingesetzt und im Rahmen des **Datenhandlings** die Vorverarbeitung, Speicherung und in vielen Fällen eine sofortige Rückmeldung an die Datenerfassungsfunktion in Form einer **Anpassung** erfolgt. Um aus den Daten **Insights** zu erstellen, sind weitere **Datenauswertungen** notwendig. Mit Hilfe von Insights können Daten auf eine neue Ebene gehoben und zur weiteren Analyse wieder vorgelegt werden. Dieser Prozess ermöglicht das Entstehen von **wirtschaftlich verwertbaren Informationen**. Im Gegensatz zum verbreiteten Narrativ wird auf dieser Ebene deutlich, dass Daten – unabhängig von ihrer verfügbaren Menge – wenig bis gar keinen wirtschaftlichen Wert haben. Tatsächlich müssen Unternehmen erhebliche Anstrengungen unternehmen, um Daten in **wirtschaftlich verwertbare Informationen** umzuwandeln.

Auf der **Kapitalschicht** wird die Rolle des Endgeräts beleuchtet, da es den Ursprung und die Quelle des analogen Signals enthält und den ersten Filter darstellt, durch den die Informationen schließlich ihren Weg in die digitalen Datensätze finden. Auf diese Daten können die **Organisationen, die die Daten erheben**, zugreifen: diejenigen, die die Kontrolle über das Gerät und/oder die auf dem Gerät ausgeführte Software und den Zugriff auf die Quelle und (in gewissem Umfang) Kontrolle darüber haben, welche Daten auf welche Weise erfasst werden. Diese Organisationen können entscheiden, ob sie diese Daten **Dritten** zur Verfügung stellen oder sie für sich selbst behalten wollen (**Exklusivität**). Der Zugriff Dritter auf die Daten erfolgt durch **Datenaustausch**. Die wichtigste Erkenntnis, die hier in der Diskussion zum Tragen kommt, ist, dass Unternehmen ihre Kontrolle über die Datenerfassung gegen die potenziellen Kostenvorteile abwägen müssen, die sich daraus ergeben, dass sie die Prozesse zur Datenerfassung und -verarbeitung nicht selbst entwickeln müssen.

Entsprechend veranschaulicht die Kapitalschicht, dass die Datenerhebung nie unabhängig ist vom jeweiligen geschäftlichen Kontext und dem Verwendungszweck. Vor allem schließt der vorgesehene Zweck nicht aus, dass die Daten in anderen Bereichen verwendet werden können. Der Zweck ist entscheidend für die Datenqualität und beeinflusst die **Genauigkeit, Vollständigkeit, Konsistenz, Zugriffsmöglichkeit und Aktualität** der Daten.

Datenökonomie – Ein bedeutender Bereich des digitalen Binnenmarkts

Datenökonomie ist kein aufstrebender Wirtschaftszweig, sondern vielmehr eine neue Dimension von Innovation und Wettbewerb, die die etablierten Geschäftsprozesse in allen Branchen in Frage stellt. Während Daten diese neue Dynamik der wirtschaftlichen Aktivität anheizen, **bringen mehr Daten nicht unbedingt mehr Wert**. Im Gegensatz zu Öl oder einer Währung sind Daten nicht-rival und nicht-materiell. Ebenso wenig äh-

nelt die Wertschöpfung aus Daten den traditionellen Wertschöpfungsketten – stattdessen findet ein kreisförmiger Wirtschaftsaustausch statt, der mehrseitige Märkte ermöglicht.

1 Einleitung

Aktuelle Innovationen ermöglichen es, (digitalisierte) Daten in einem noch nie dagewesenen Umfang zu erfassen, zu analysieren und zu speichern. Auch wenn unterschiedliche Quellen zu unterschiedlichen spezifischen Zahlen kommen¹, ist es unbestreitbar, dass mit der so genannten "Datenökonomie" ein enormes wirtschaftliches Potenzial verbunden ist.

Digitalisierung und digitale Daten betreffen alle Bereiche der Wirtschaft; es ist schwer vorstellbar, dass langfristig nicht jeder Bereich seine Anstrengungen zur Nutzung digitaler Daten irgendwie verstärken muss. Während Daten ebenso vielfältig sein können wie die Branchen selbst, scheint die allgemeine Debatte über die Nutzung personenbezogener Daten vor allem durch den Aspekt der Monetarisierung der im Internet angebotenen Dienstleistungen durch gezielte Werbung bestimmt zu sein. Unser Bericht versucht, den Horizont der Diskussion zu erweitern und zum besseren Verständnis des Sachverhalts beizutragen.

In **Kapitel 2** beleuchten wir einige der **grundlegenden Eigenschaften von Daten**. Dabei untersuchen wir einige der potenziell überzogenen Erwartungen und Annahmen bezüglich der Fähigkeit von Daten, wirtschaftliche Aktivitäten zu erleichtern. Unser Bericht untersucht die Rolle der Datenqualität neben der Datenquantität, die normalerweise im Mittelpunkt des Interesses steht.

Kapitel 3 untersucht, wie die Datenökonomie funktioniert, und erläutert die Rolle von Daten in digitalen Geschäftsmodellen, insbesondere digitalen Plattformen. Daten spielen eine wesentliche Rolle bei der Innovation; bisher scheint jedoch unklar zu sein, welche Faktoren die Rolle von Daten bei der Innovation beeinflussen und ob ein fehlender Datenzugang tatsächlich Innovationen behindert.

Kapitel 4 konzentriert sich auf die Rolle von Daten für den Wettbewerb. Es wird über Daten als Input und als wichtige Ressource im Wettbewerb diskutiert, insbesondere im Hinblick auf die Datenqualität. Dieses Kapitel betrachtet die starken wirtschaftlichen Einflüsse, die der Struktur der Datenökonomie zugrunde liegen, und das Potenzial für datengetriebene Marktmacht und Marktzutrittsschranken. Im Anschluss werden Maßnahmen zur Förderung eines wirksamen Wettbewerbs in der Datenökonomie vorgeschlagen.

¹ Zum Beispiel haben vor einigen Jahren Manyika J, Chui M, Groves P, Steve F, Kuiken V, Doshi EA. 2013. Open data: Unlocking innovation and performance with liquid information, McKinsey Global Institute, den jährlichen Wert, der durch offene Daten in sieben verschiedenen Bereichen ermöglicht wird, mit rund 3 Billionen Dollar weltweit bewertet. IDC, Open Evidence. 2017. European Data Market SMART 2013/0063 - Final Report. A study prepared for the European Commission, IDC, Open Evidence, schätzt den Wert der Data Economy in Europa für das Jahr 2016 auf etwa 300 Milliarden Euro. Für die Zukunft, so Hogan O, Holdgate L, Jayasuriya R. 2016. The Value of Big Data and the Internet of Things to the UK Economy, Cebr, London, wird der Beitrag von Big Data sowie des IoT bis 2020 allein im Vereinigten Königreich 62 Milliarden Pfund erreichen.

In Kapitel 5 werden mehrere Ansätze für den Datenaustausch beleuchtet und ihre jeweiligen Vor- und Nachteile erläutert.

Kapitel 6 greift auf die im Bericht gewonnenen Erkenntnisse zurück, um eine Zusammenfassung der **Herausforderungen der Datenökonomie** zu präsentieren und die **Schlussfolgerungen auf politischer Ebene** für Entscheidungsträger sowie Wettbewerbs- und Regulierungsbehörden zusammenzufassen und einen Ansatz vorzuschlagen, wie durch die Verwendung von **Referenzarchitekturen** ein **intensiverer und verbesserter Datenaustausch** erleichtert werden kann.

Kapitel 7 schließt den Bericht ab.

Die wichtigsten Ergebnisse aus Kapitel 2

- *Um den wirtschaftlichen Wert von Daten zu verstehen, ist es entscheidend, zu erkennen, dass Daten nicht als objektive Fakten betrachtet werden können, sondern als bewusst ausgewählte Darstellungen der Realität aus dem Kontext und dem Zweck, für den sie erfasst wurden.*
- *Die Verarbeitung kann den wirtschaftlichen Wert von Daten in einem Kontext erhöhen, in anderen jedoch zunichte machen.*
- *Datenmenge und Datenqualität sind eng miteinander verknüpft. Um den wirtschaftlichen Wert von Daten zu erkennen, kann das eine nicht ohne das andere betrachtet werden.*
- *Daten sind immateriell und nicht-konkurrierend. Vor allem aber unterscheiden sie sich nicht nur in Form und Format, sondern auch in Bezug auf Kontext und Zweck.*
- *Ein bestimmtes Datum kann auf viele Arten erfasst und zugänglich gemacht werden. Echte Exklusivität des Datenzugangs ist in einer zunehmend digitalisierten Welt nur schwer herzustellen.*
- *Es gibt eine abnehmende Rendite, wenn die Datenmenge vergrößert wird. In den meisten Kontexten ist jedoch eine Mindestdatenmenge erforderlich. Die Größe dieser Mindestmenge hängt stark vom Kontext und Zweck der Datennutzung ab.*
- *Daten sind kein Wundermittel. Theorie, Wissen und Fähigkeiten sind (mindestens) ebenso wichtige Faktoren für den wirtschaftlichen Wert von Daten.*
- *Datentypologien können nur einen Teil des Prozesses von analogen Signalen über digitale Daten bis hin zu wirtschaftlich verwertbaren Informationen abdecken. Für ein umfassendes Verständnis der Faktoren, die Daten und Datenwert beeinflussen, ist ein ganzheitlicher Ansatz erforderlich.*
- *Um Daten effektiv zu typologisieren, ist es unerlässlich, den spezifischen Kontext und den vorgesehenen Verwendungszweck der Daten zu beachten. Generell ist es hilfreich, zwischen 1) strukturierten Daten, 2) unstrukturierten, sich wiederholenden Daten und 3) unstrukturierten, sich nicht wiederholenden Daten zu unterscheiden.*

2 Daten richtig nutzen

Mit dem Internet und einer steigenden Anzahl damit verbundener Endgeräte hat sich unsere Fähigkeit erhöht, analoge Signale in digitale Bits zu übertragen (d.h. unsere Fähigkeit zur Digitalisierung). Wenn wir von Daten sprechen, denken wir typischerweise an digitale Bits, die Informationen speichern und sie für die Auswertung zugänglich machen, was wiederum die Verwendung der Informationen in verschiedenen Kontexten ermöglicht. Dieser Prozess ist das Herzstück der digitalen Transformation, die Unternehmen und unsere Gesellschaft durchlaufen. Erhebliche Verbesserungen in der Datenspeicherung und -analyse haben innovative Anwendungen digitaler Daten ermöglicht. Zu Beginn unserer Diskussion über die Datenökonomie stellen wir kurz ein Referenzmodell vor, das wir auf der Grundlage unserer Untersuchung als integrierten Ansatz zur Identifizierung der wichtigsten Herausforderungen empfehlen, auf die in späteren Abschnitten ausführlich eingegangen wird. Unser Referenzmodell wird im folgenden Abschnitt beschrieben.

2.1 Mehrschichtiges Referenzmodell

***Erkenntnisse:** Nur ein ganzheitliches Verständnis des Prozesses – des Übergangs von analogen zu digitalen Daten hin zur Generierung von wirtschaftlich verwertbaren Informationen – kann die Diskussion über die Datenökonomie wirklich bereichern. Eine einheitliche Datentypologie ist den allermeisten Fällen irreführend und zwangsläufig unvollständig.*

Unsere Fähigkeit, Daten zu erfassen und zu nutzen, hat in den letzten Jahrzehnten dramatisch zugenommen, hauptsächlich mit digitalen Technologien, die Systeme, Geräte und Ressourcen umfassen, die digitale Daten erzeugen, speichern, verarbeiten, austauschen oder nutzen. Der Prozess der Umwandlung eines analogen Signals, das Informationen (z.B. Ton, Bild, gedruckter Text) übermittelt, in binäre Bits wird im anglo-amerikanischen Sprachraum als „**Digitization**“ bezeichnet. Die Anwendung oder verstärkte Nutzung digitaler Technologien durch eine Organisation, eine Branche oder ein Land, durch die bestehende Aufgaben transformiert oder neue ermöglicht werden, wird als **Digitalisierung** bezeichnet. Dieser Begriff bezieht sich auf die Auswirkungen der *Digitization* auf die Wirtschaft oder Gesellschaft (OECD & Eurostat 2018). Im Rahmen dieses Berichts folgen wir dieser Terminologie. Demnach ist die *Digitization* Vorbedingung für die Datenökonomie.

Unser Referenzmodell soll die (wirtschaftlichen) Eigenschaften von Daten veranschaulichen und besteht aus vier Schichten. Die erste Schicht stellt die Umwandlung von **analogen Daten** (Signale aus der realen Welt) in **digitale Daten** dar. Auf dieser **Datenschicht** können die digitalen Daten weiter in **strukturierte, unstrukturierte, sich wiederholende und unstrukturierte, sich nicht wiederholende** digitale Daten unter-

teilt werden. Während andere Arten der Darstellung je nach Kontext nützlich sein können, geben sie doch durchweg keinen Aufschluss über die Datenbeschaffenheit und ihre Auswirkungen auf die Datennutzung. Dies ist aber besonders wichtig, da die große Mehrheit der heute verfügbaren digitalen Daten unstrukturiert ist (z.B. Video, Bilder, Text, natürliche Sprache, Social Media Streams).

Die **Prozessschicht** macht deutlich, dass es keine digitalen Rohdaten gibt. Bereits bei der Digitalisierung wurde bewusst entschieden, welche analogen Signale in digitale Bits umgewandelt werden sollen und wie genau diese Umwandlung erfolgt, z.B. welche Sensoren oder Schnittstellen eingesetzt werden. Der Verarbeitungsgrad steigt in der Regel mit der weiteren Nutzung der digitalen Daten.

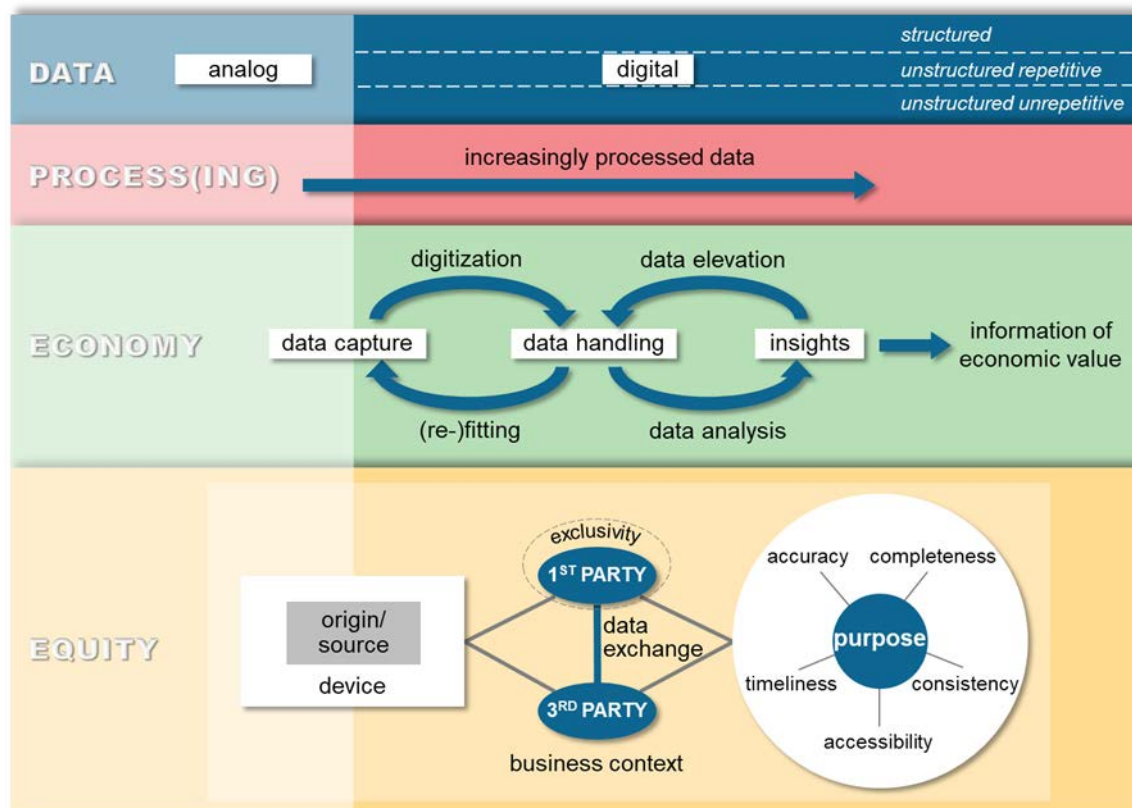
Die **Wirtschaftsschicht** zeigt, dass die Umwandlung von Daten in wirtschaftlich verwertbare Informationen nicht zwangsläufig ein einfacher Prozess ist. Erst in Kombination mit einem spezifischen Kontext können Daten zu wirtschaftlich verwertbaren Informationen werden. Logischerweise erfolgt die Digitalisierung während der **Datenerfassung**, wobei Geräte mit Sensoren und/oder Schnittstellen eingesetzt und im Rahmen des **Datenhandlings** die Vorverarbeitung, Speicherung und in vielen Fällen eine sofortige Rückmeldung an die Datenerfassungsfunktion in Form einer **Anpassung** erfolgt. Um aus den Daten **Insights** zu erstellen, sind weitere **Datenanalysen** notwendig. Mit Hilfe von Insights können Daten auf eine neue Ebene gehoben und zur weiteren Auswertung wieder vorgelegt werden. Dieser Prozess ermöglicht das Entstehen von **wirtschaftlich verwertbaren Informationen**. Im Gegensatz zu der verbreiteten Annahme wird auf dieser Ebene deutlich, dass Daten – unabhängig von ihrer verfügbaren Menge – wenig bis gar keinen wirtschaftlichen Wert haben. Tatsächlich müssen Unternehmen erhebliche Anstrengungen unternehmen, um die Daten in wirtschaftlich verwertbare Informationen umzuwandeln.

Auf der **Kapitalschicht** wird die Rolle des **Endgeräts** beleuchtet, da es den **Ursprung und die Quelle** des analogen Signals enthält und den ersten Filter darstellt, durch den die Informationen schließlich ihren Weg in die digitalen Datensätze finden. Auf diese Daten können die **First-Party-Organisationen** zugreifen: diejenigen, die die Kontrolle über das Gerät und/oder die auf dem Gerät ausgeführte Software und den Zugriff auf die Quelle und (in gewissem Umfang) Kontrolle darüber haben, welche Daten auf welche Weise erfasst werden. Diese First-Party-Organisationen können entscheiden, ob sie diese Daten **Dritten** zur Verfügung stellen oder sie für sich selbst behalten wollen (**Exklusivität**). Der Zugriff Dritter auf die Daten erfolgt durch **Datenaustausch**.

Entsprechend veranschaulicht die Kapitalschicht, dass die Datenerhebung nie unabhängig ist vom **Geschäftskontext** und dem **Zweck** der Datennutzung. Insbesondere schließt der jeweilige Zweck nicht aus, dass die Daten auch in anderen Zusammenhängen verwendet werden können. Durch die Anpassung der Datenverarbeitung an einen bestimmten Zweck kann jedoch die Verwertbarkeit der Daten für andere Zwecke beeinträchtigt werden. Der Zweck ist entscheidend für die Datenqualität und hat Einfluss auf

die **Genauigkeit, Vollständigkeit, Konsistenz, Zugänglichkeit** und **Aktualität** der Daten. Insbesondere kann der Kontext zwischen "datenreichen" und "datenarmen" Umgebungen in Abhängigkeit vom Grad der Digitalisierung in diesem speziellen Kontext variieren.

Abbildung 2-1: Referenzmodell zur Darstellung der (wirtschaftlichen) Eigenschaften von Daten



Quelle: WIK-Consult.

Zusammengefasst zeigt unser Referenzmodell verschiedene Unzulänglichkeiten und Fehlschlüsse der verbreiteten Narrative rund um die Datenökonomie auf.

- "Datenlieferanten" und "Datennutzer" können nicht auf eine einfache und eindeutige Weise unterschieden werden, wie es das DataLandscape-Projekt² vorschlägt. Da die Datenerfassung ein wichtiger Bereich in praktisch allen Organisationen geworden ist, wird jede Organisation bei einigen Datenerfassungs- oder Datenerhebungsvorgängen über den direkten Zugriff verfügen, während für nicht selbst erfasste oder erhobene Daten nur der Zugriff als Dritter möglich ist.

² <http://datalandscape.eu/>

Die Webseite sowie die dort präsentierten Berichte und Daten sind Teil der EU-Datenmarktstudie, die von der Europäischen Kommission im Rahmen von SMART 2013/0063 in Auftrag gegeben wurde.

- Die Heterogenität von Daten ist nicht nur darauf zurückzuführen, dass sie aus verschiedenen Quellen stammen, sondern sie ist eine Kombination aus Unterschieden in Bezug auf Kontext, Erhebungszweck, Quelle, Endgerät und Datenverarbeitung. Dieser Umfang an Heterogenität wird in anderen Studien oft vernachlässigt. (z. B. Arnaut et al (2018), Crémer et al (2019), Furman et al (2019), Morton et al (2019) Feld (2019)).
- Im Gegensatz zur Auffassung verschiedener Studien ist die Nutzung von Big Data kein Selbstläufer, der wirtschaftlichen Erfolg in der Datenökonomie ermöglicht. Vielmehr stellt sie eine komplexe Aufgabe dar, die richtige Balance zwischen Datenqualität und Datenmenge zu finden. Je nach Kontext und Zweck der Datennutzung müssen Unternehmen mehrere Dimensionen der Datenqualität wie Genauigkeit, Vollständigkeit und Aktualität berücksichtigen. Es gibt zwar eine mindestoptimale Datenmenge, aber es gibt auch abnehmende Grenzerträge bei zunehmender Datensatzgröße. Daher müssen Unternehmen, die Big Data verwenden, mehrere Abwägungen treffen, wobei die Bedeutung von einschlägigem Fachwissen (z.B. Data Scientist) und Erfahrung (z.B. Learning-by-doing Feedback-Schleifen) deutlich wird.
- Die Diskussion über die Relevanz von Daten scheint davon auszugehen, dass Daten an sich einen Wert haben. Daten als solche sind jedoch wertlos. Tatsächlich ist es für eine fundierte Diskussion über die Datenökonomie sehr wichtig, zu verstehen und zu erkennen, dass nur Daten in Verbindung mit dem Kontext zu Informationen werden, die wiederum Wertschöpfung ermöglichen.

2.2 Definition des Begriffs “Daten”

***Erkenntnisse:** Erst Daten in Verbindung mit dem Kontext werden zu Informationen, die wiederum Wertschöpfung ermöglichen. Die Wiederverwendung von Daten hängt vom Wissen über den Kontext und den Zweck ab, für den die Daten ursprünglich erfasst wurden.*

Trotz seiner offensichtlich zunehmenden Bedeutung haben der Begriff "Daten" und seine Definition bei Wissenschaftlern nur wenig Aufmerksamkeit erhalten. Überraschenderweise haben akademische Disziplinen wie Rechts- und Wirtschaftswissenschaften praktisch keine Anstrengungen unternommen, die Eigenschaften von Daten wirklich zu ergründen, während sie erhebliche Anstrengungen in die Erforschung und Diskussion der Auswirkungen von Daten auf Wohlstand, Wirtschaftswachstum und Wettbewerb investiert haben. Wie Furner (2016) feststellte: *“Eine Quelle von Missverständnissen in den heutigen Diskussionen über Datenwissenschaft und Big Data ist die Tendenz, drei verwandte, aber unterschiedliche Interpretationen zu vermischen: Daten als Fakten, Daten als typischerweise numerische Attributwerte und Daten als Bits”*. (S. 298) Deshalb beginnt unser Beitrag damit, die spezifische Definition des Begriffs, an den wir uns halten, zu präzisieren.

Für den vorliegenden Beitrag verwenden wir das Wort „*datum*“ für den einzelnen Datenpunkt. *Daten* ist der Begriff, mit dem wir eine Vielzahl von einzelnen Datenpunkten bezeichnen. Was die Definition von Daten als solche betrifft, stimmen wir mit der Definition überein, die von Kaase (2001) vorgeschlagen und von Hjørland (2019) auf der Grundlage seiner Bestandsaufnahme der aktuellen Definitionen des Begriffs als die geeignetste vorgeschlagen wurde:

“*Daten sind Informationen über Eigenschaften von Analyseeinheiten.*”

Diese Definition scheint anderen Definitionen überlegen zu sein, da sie Daten nicht mit Dokumenten verwechselt.³ Sie umfasst die in den Daten erfassten Einheiten-Phänomene, die eine notwendige Voraussetzung für die Interpretierbarkeit der Daten sind.⁴ Dies ist auch deshalb wichtig, weil die Definition berücksichtigt, dass die Wahl, das Verständnis und die Beschreibung einer "Einheit" vom Kontext der Datenerfassung, -erhebung oder -interpretation abhängen. Am wichtigsten ist, dass diese Definition anerkennt, dass Daten nicht für sich selbst sprechen, da sie keine objektiven Fakten sind (Hjørland 2019). Tatsächlich werden nur Daten in Kombination mit dem Kontext zu Informationen, die wiederum eine Wertschöpfung ermöglichen. Allerdings empfiehlt es sich, zu untersuchen, wie sich das Verständnis von Daten im Laufe der Zeit entwickelt hat, um einige der wichtigsten Schwächen der aktuellen Diskussion um die so genannte Datenökonomie zu beleuchten.

Der Begriff "Daten" geht auf das Lateinische zurück.⁵ *Datum* (d.h. der Singular) im wörtlichen Sinne des Wortes bedeutet "das, was gegeben wird" und *Daten* (d.h. der Plural) "gegebene Dinge" oder einfach "Gaben". Dieser Ursprung prägte die Art und Weise, wie frühe Mathematiker wie Euklid den Begriff "Daten" verwendeten, um sich auf gegebene Fakten innerhalb einer Gleichung zu beziehen, wie z.B. die Länge der Seiten und Winkel innerhalb von Dreiecken, von denen einige gegeben sein können, während andere nach mathematischen Prinzipien berechnet werden können. Dieses Verständnis von Daten hielt bis Mitte des 19. Jahrhunderts an, als z.B. Worcester's Wörterbuch der englischen Sprache Daten als "*Tatsachen oder Prämissen, die gegeben oder angenommen werden, um daraus Schlussfolgerungen abzuleiten, die Tatsachen, aus denen eine Schlussfolgerung gezogen wird*" definiert. Dieses Verständnis verlagerte sich jedoch in der zweiten Hälfte des 19. Jahrhunderts, als mit den aufkommenden Disziplinen der Statistik- und Sozialwissenschaften die Zahl der Tafeln zunahm, die eine systema-

³ Für eine Erläuterung des Unterschieds zwischen Daten und Dokumenten siehe Furner J. 2016. "Data": The data In *Information Cultures in the Digital Age - A Festschrift in Honor of Rafael Capurro*, ed. M Kelly, J Bielby, pp. 287-306. Wiesbaden: Springer.

⁴ Für weitere Erläuterungen hierzu siehe Jensen HE. 1950. Editorial Note In *Through Values to Social Interpretation: Essays on Social Contexts, Actions, Types and Prospects*, ed. H Becker, pp. vii-xi. Durham, NC: Duke University Press cf. Hjørland B. 2019. Data (with Big Data and Database Semantics). *KO Knowledge Organization* 45: 685-708.

⁵ Konkret stammt der Begriff aus dem Präsens Indikativ des Verbs *dō* ("ich gebe"), dessen Partizip Perfekt *datus* ("gegeben") ist. Die beiden heute am häufigsten verwendeten Formen *datum* und *data* sind der Nominativ des jeweiligen Partizips, der Nominativ des Neutrum-Singulars und der Nominativ des Neutrum-Plurals sowie der Nominativ des Femininum-Singulars.

tisch organisierte Aufzeichnung und Übermittlung von auf Beobachtungen und Messungen beruhenden Häufigkeiten und Größen ermöglichten. Plötzlich wurden diese einmal gesammelten und in schriftlicher Form fixierten Datentafeln als "gegeben" behandelt und wurden so zum Rohmaterial für neuartige Formen der quantitativen Analyse. Die Einträge wurden schließlich als Daten bezeichnet (Furner 2016).⁶

Die Behandlung solcher auf Beobachtungen und Messungen beruhender Häufigkeiten und Größen als "gegeben" oder sogar als Tatsachen ist ein wesentlicher Fehlschluss, der in der naiv-realistischen ontologischen Annahme wurzelt, dass Phänomene unabhängig von den jeweiligen Beobachtern existieren und Daten völlig objektiv aus einer einzigen objektiven Realität abgelesen werden können. Wie Drucker (2011) aufzeigt, vernachlässigt ein solches Verständnis von Daten den "situativen, partiellen und konstitutiven" Charakter der Erkenntnisgewinnung völlig. Folglich schlägt sie vor, Daten durch *capta* zu ersetzen, d.h. das „Gegebene“ durch das „Genommene“ im wörtlichen Sinne zu ersetzen. Checkland (1999) und Capurro (1978), wie sie in Zins (2007) zitiert werden, teilen diese Ansicht.

Hjørland (2019) betont in diesem Zusammenhang, dass *"[...] dokumentierte Daten für die Zwecke, für die sie erhoben werden, als Fakten angesehen werden; sie stellen ausreichende Fakten in einem gegebenen sozialen Kontext dar. Wäre dies nicht der Fall, wäre es unmöglich, auf der Grundlage von Daten zu handeln; so wäre es beispielsweise unmöglich, Stammbäume zu erstellen, und es gäbe keinen Grund, solche Dokumente (z. B. Geburtsurkunden) auszustellen. [...] Das Datum "X ist das Kind von Y" kann z.B. aus drei Arten von Dokumenten gewonnen werden:*

- (1) *einem Interview mit Y;*
- (2) *der Geburtsurkunde von X;*
- (3) *und der DNA-Analyse basierend auf der DNA von X und Y.*

Normalerweise wird davon ausgegangen, dass alle drei Dokumente dieses Datum als Tatsache wiedergeben. Der DNA-Bericht ist heute die zuverlässigste Quelle, aber alle drei enthalten das Datum" (S. 688).

All diese Autoren weisen auf ein offensichtliches Manko der aktuellen Diskussion um die Datenökonomie hin, die von der impliziten Annahme getrieben zu sein scheint, dass Daten eigenständige, objektive Fakten oder Beobachtungen sind, obwohl sie in Wirklichkeit keines von beiden sind und sein können, oder, wie Gitelman (2013) es ausdrückte: *"'Rohdaten' sind ein Widerspruch".⁷*

⁶ Furner J. 2016. "Data": The data. In *Information Cultures in the Digital Age - A Festschrift in Honor of Rafael Capurro*, ed. M Kelly, J Bielby, pp. 287-306. Wiesbaden: Springer bietet eine sehr viel detailliertere und aufschlussreichere Geschichte der Begriffe *Daten* und *Datum* über die Jahrhunderte hinweg. Sein Buchkapitel ist hier nur sehr kurz zusammengefasst.

⁷ Etwas kontraintuitiv erscheint Big Data - aufgrund einer implizit angenommenen Objektivität der Sensoren selbst als eine höchst unmittelbare Erfassung realer Handlungen menschlicher und nicht-

Das Verständnis der situativen und zweckentsprechenden Definition und Erhebung von Daten ist für die Datenökonomie von wesentlicher Bedeutung. Zum einen steht dies im Gegensatz zu der von manchen vertretenen Auffassung, dass Daten lediglich ein Nebenprodukt der Digitalisierung sind.⁸ Zweitens kann das Verständnis des grundlegenden Charakters von Daten die aufkommende Debatte über den Zugang zu und die Wiederverwendung von Daten, z.B. in Form von Datenaustausch, maßgeblich beeinflussen. Wie Borgman (2010) feststellt: *“Allzu selten definieren diejenigen, die die gemeinsame Nutzung und Pflege von Daten fördern, “Daten” ausdrücklich oder berücksichtigen die Vielfalt der Formen, die Daten annehmen können.”* (S. 2).

Ein großer Teil der Erwartungen, die mit der Datenökonomie und dem sozioökonomischen Potenzial des Austauschs oder anderweitigen Zugänglichmachens von Daten für die (Wieder-)Verwendbarkeit verbunden sind, scheint auf dem Konzept von Big Data und dem so genannten "vierten Paradigma" zu beruhen⁹, das durch Big Data ermöglicht wird.¹⁰ Im Kern verkündet das vierte Paradigma ein neues Zeitalter der Erkenntnis auf der Grundlage von Big Data, in dem die Theorie nicht mehr der Ausgangspunkt für die Suche nach Wissen ist, sondern die Daten als solche diese Rolle übernehmen (Dhar 2013). Ziel ist es, *“interessante und dauerhafte Strukturen zu identifizieren, die die Daten bestätigen”* und die in der Zukunft zu erwarten sind (Dhar 2013). Es wird erwartet, dass die auf der Grundlage solcher Muster gewonnenen Erkenntnisse ein enormes Potenzial für Unternehmen bei der Gründung neuer Unternehmen, der Ent-

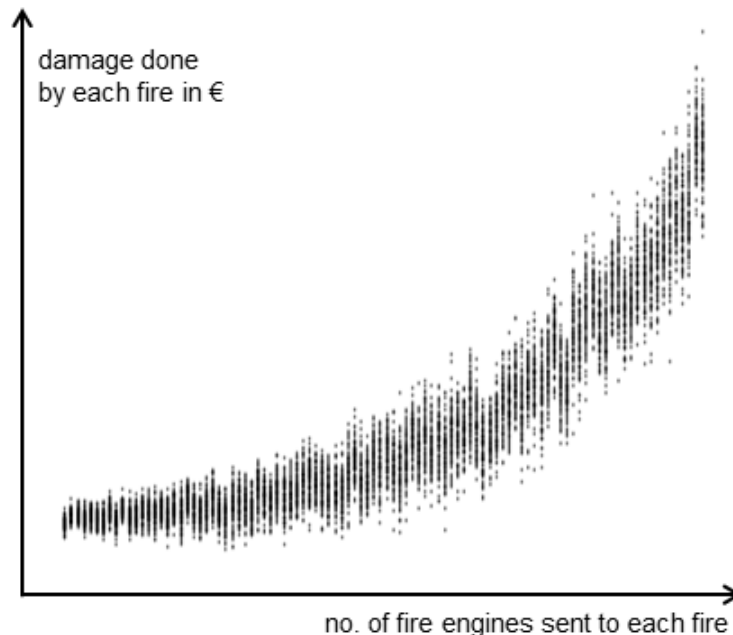
menschlicher Art betrachtet - besonders anfällig für die hier skizzierten erkenntnistheoretischen Herausforderungen. Vor allem große Datenströme werden in der Regel einer automatischen Qualitätsprüfung unterzogen, deren Hauptzweck darin besteht, "Ausreißer" zu identifizieren und zu korrigieren, die aus den im konkreten Fall verwendeten Datenquellen stammen. Es gibt zwar verschiedene Wege, diese Korrektur zu realisieren, aber allen diesen Wegen ist gemeinsam, dass sie eine Art Vorhersage auf der Grundlage historischer Daten aus derselben Quelle entwickeln, mit denen jedes neue *Datum* verglichen wird, und wenn die Abweichung einen vordefinierten Bereich überschreitet, wird das *Datum* normalerweise durch ein *Datum* ersetzt, das den historischen Daten aus dieser Quelle besser entspricht, z.B. das entsprechende Minimum, Maximum oder den Median. Alternativ dazu wird das *Datum* völlig außer Acht gelassen. Anstatt die Eingabe als objektiv zu behandeln, wird sie also immer automatisch überprüft und kann korrigiert werden. Einige dieser Systeme beruhen auf Deep Learning-Algorithmen, die sogar das Verfahren, wie und warum ein bestimmtes *Datum* ersetzt wurde, für menschliche Akteure undurchsichtig machen können. Zweitens macht die Heterogenität von Big Data diese anfällig für erkenntnistheoretische Herausforderungen, die sich aus inkonsistenten oder sogar schlichtweg widersprüchlichen Dokumenten, die sich auf ein und dasselbe *Datum* beziehen, ergeben. Auch die Art und Weise, wie die Auswahl getroffen wird, welches dieser Dokumente das wahre *Datum* enthält (und damit Teil des Datensatzes wird), ist in der Regel undurchsichtig. Mirzaie M, Behkamal B, Paydar S. 2019. Big Data Quality: A systematic literature review and future research directions. *arXiv preprint arXiv:1904.05353*.

- 8 Ob Daten als Nebenprodukt betrachtet werden können oder nicht, kann von der Perspektive abhängen. Aus der Perspektive eines Datensubjekts – das eine Person oder eine Organisation sein kann – können Daten ein Nebenprodukt in dem Sinne sein, dass das Datensubjekt wenig oder keine Kontrolle über die im Rahmen der von ihm durchgeführten Aktionen erfassten Daten hat. Aus der Perspektive des Unternehmens, das die Daten erfasst und möglicherweise speichert und nutzt, ist eine aktive Entscheidung erforderlich, um dies zu tun. Daher können solche Daten nicht als Nebenprodukt der Datenwertschöpfung betrachtet werden. Hier ist die zielgerichtete und kontextualisierte Erfassung von Daten immer der Ausgangspunkt. Dies schließt jedoch nicht aus, dass einmal erfasste Daten für andere als die ursprünglich vorgesehenen Zwecke verwendet werden.
- 9 Der Begriff des "vierten Paradigmas" wurde ursprünglich von Microsoft Research geprägt. Vgl. Hey T, Tansley S, Tolle K. 2009. *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Redmont, WA: Microsoft Research.
- 10 Zur Diskussion über Datenqualität und -quantität siehe Abschnitt 2.5.

wicklung neuer Dienstleistungen und Produkte und der Verbesserung von Geschäftsabläufen haben werden (Lee 2017). Cao (2017) ist bei der Definition von Datenprodukten etwas vorsichtiger: *"Ein Datenprodukt ist ein Ergebnis von Daten oder wird durch Daten ermöglicht oder getrieben und kann eine Entdeckung, eine Vorhersage, eine Dienstleistung, eine Empfehlung, ein Einfluss auf die Entscheidungsfindung, ein Denken, ein Modell, ein Modus, ein Paradigma, ein Werkzeug oder ein System sein. Die ultimativen Datenprodukte von Wert sind Wissen, Intelligenz, Weisheit und Entscheidung"* (S. 51).

Beispiel

Abbildung 2-2: Beispiel für eine potentiell irreführende Korrelation



Die oben genannten Daten über viele Brände zeigen eine offensichtlich positive Korrelation zwischen dem Schaden, der durch einen Brand verursacht wurde, und der Anzahl der Löschfahrzeuge, die zu diesem Brand geschickt wurden.

Allein aufgrund der Daten müsste man zu der Schlussfolgerung kommen, dass es notwendig ist, weniger Löschfahrzeuge zu einem Brand zu schicken, um den Schaden zu reduzieren. Natürlich ist das Gegenteil der Fall. Ohne theoretische Kenntnisse über die Kausalzusammenhänge der beobachteten Phänomene kann man also sehr leicht zu falschen Schlussfolgerungen gelangen. Folglich sollte man vorsichtig sein, Muster in den Daten als eindeutig zu betrachten.

Sowohl aus erkenntnistheoretischer als auch aus praktischer datenwissenschaftlicher Sicht wird die Prämisse, dass Daten die Theorie ersetzen können, als falsch angesehen. Boyd and Crawford (2012) integrieren ihre Kritik an diesem Konzept sogar in ihre Definition von Big Data, wenn sie es als ein *"kulturelles, technologisches und wissenschaftliches Phänomen"* behandeln, das auf der Interaktion folgender Faktoren beruht:

- (1) *Technologie: Maximierung der Rechenleistung und der algorithmischen Genauigkeit, um große Datensätze zu sammeln, zu analysieren, zu verknüpfen und zu vergleichen.*

(2) *Analyse: Nutzung großer Datensätze zur Identifizierung von Mustern, um wirtschaftliche, soziale, technische und rechtliche Aussagen zu treffen.*

(3) *Mythologie: die weit verbreitete Annahme, dass große Datensätze eine höhere Form von Intelligenz und Wissen bieten, die mit der Aura von Wahrheit, Objektivität und Genauigkeit Einsichten erzeugen können, die vorher unmöglich waren" (S. 663).*

Viele andere Forscher teilen diese Kritik. So weist Frické (2009) in seiner Kritik an der Hierarchie „Daten-Information-Wissen-Weisheit“ (DIKW) darauf hin, dass die Vorstellung, Daten seien (automatisch) die Quelle von Informationen und Informationen beantworteten Fragen, das willkürliche und sinnlose Sammeln von Daten fördert. Dieses Datensammeln wird von der Erwartung angetrieben, dass diese Daten zu Informationen avancieren, was vielleicht nie geschehen wird.¹¹ Hjørland (2019) stellt fest, dass das *"vierte Paradigma ... ein Hemmnis für ein besseres Verständnis theoretischer Probleme im Zusammenhang mit der Daten- und Wissensorganisation darstellt. Deshalb – obwohl Big Data der Ursprung von 'e-science' ist – wird Big Data nicht durch e-science definiert und kann nicht ohne eine entsprechende Theorie auskommen"* (S. 701). Tatsächlich scheint ein Großteil der Aufmerksamkeit für Big Data in einem problematischen (naiven) Empirismus zu wurzeln.¹² Die Kritik an diesem Empirismus entwertet jedoch nicht die Daten oder die Datenökonomie; sie unterstreicht nur die Notwendigkeit, die Beschaffenheit der Daten zu verstehen und ihr unbestreitbares sozioökonomisches Potenzial auf der Grundlage weiterer Annahmen zu erforschen.

Über ein grundlegendes Verständnis der Daten als solche hinaus ist es im Rahmen der Datenökonomie besonders wichtig, auch die wirtschaftlichen Merkmale von Daten zu erörtern. Dies geschieht im folgenden Abschnitt.

2.3 Charakterisierung von Daten aus wirtschaftlicher Sicht

Erkenntnisse: *Daten werden nicht-rival verwendet, sie können vertraulich behandelt (Exklusivität der betreffenden Organisation) oder öffentlich gemacht werden (z.B. durch Veröffentlichung im Internet) und ihr Wert hängt von der Aktualität, dem Kontext und dem beabsichtigten Verwendungszweck ab.*

¹¹ Folglich sind die Konzepte von Big Data und Data Mining häufig Gegenstand kontroverser Diskussionen in der Informations- und Datenwissenschaft. Vgl. z. B. Austin PC, Goldwasser MA. 2008. Pisces did not have Increased Heart Failure: Data-driven Comparisons of Binary Proportions between Levels of a Categorical Variable can Result in Incorrect Statistical Significance Levels. *Journal of Clinical Epidemiology* 61: 295-300, Austin PC, Mamdani MM, Juurlink DN, Hux JE. 2006. Testing Multiple Statistical Hypotheses Resulted in Spurious Associations: a Study of Astrological Signs and Health. *Ibid.* 59: 871-72, Frické M. 2015. Big Data and its Epistemology. *Journal of the Association for Information Science and Technology* 66: 651-61.

¹² Dies wirft auch Zweifel am letztendlichen Erfolg der so genannten "Computational Theory Discovery" auf, vgl. z.B. Berente N, Seidel S, Safadi H. 2018. Research Commentary—Data-Driven Computationally Intensive Theory Development. *Information Systems Research* 30: 50-64.

Jede Disziplin scheint dem Begriff Daten und ihrer Verwendung ihre ganz eigenen Vorurteile, Kenntnisse und Traditionen zuzuschreiben. In Abschnitt 2.4 zeigen wir, dass sich dies auch aus den jeweiligen Kategorisierungen von Daten aus den verschiedenen Disziplinen ergibt. Wir gehen von einer ökonomischen Perspektive aus, ohne wichtige Erkenntnisse aus anderen Forschungsdisziplinen zu vernachlässigen.

Aus wirtschaftlicher Sicht sind Daten ein immaterielles Gut und können daher mehrfach und zu verschiedenen Zeitpunkten für unterschiedliche Zwecke genutzt werden (Floridi 2010, Hildebrandt & Arnold 2016, Schepp & Wambach 2016). Daten sind nicht-rival in ihrer Verwendung, da sie als Faktor viele Male, gleichzeitig oder nacheinander, verwendet werden können. Wenn Daten von einem Unternehmen erfasst oder erhoben werden, hindert dies andere Unternehmen nicht daran, die gleichen Daten zu erheben. Wie in dem oben genannten Beispiel von Hjørland (2019) ausgeführt, kann ein und dasselbe Datum auf verschiedene Weise erfasst werden. So können alle daran interessierten Unternehmen entweder die gleichen oder unterschiedliche Wege zur Erfassung wählen.

Eine weitere Schlüsseleigenschaft der Daten – sobald sie erfasst wurden – besteht darin, dass andere Akteure von ihrer Verwendung ausgeschlossen werden können. Einerseits können Daten in wirtschaftlicher Hinsicht ein "Clubgut" (Buchanan 1965) sein, wenn eine Konstellation von Nicht-Rivalität und Ausschließbarkeit besteht. Andererseits können Daten bei Nicht-Ausschließbarkeit, z.B. durch die Veröffentlichung im Internet (z.B. Open Access), auch zu einem "öffentlichen Gut" werden (Cornes & Sandler 1986). Daher ist die beliebte Vorstellung von Daten als Öl oder Währung trügerisch, da ihr Wert entscheidend von ihrem Kontext, ihrer Genauigkeit und Aktualität abhängt. So werden Daten im Gegensatz zu Öl beispielsweise nicht verbraucht, wenn sie verwendet werden. Da der Wert von Daten vom Kontext und vom beabsichtigten Verwendungszweck abhängt, ist es auch äußerst schwierig, einen bestimmten Preis für die Daten festzulegen. Klar ist jedoch, dass ein einzelnes Datum für sich genommen typischerweise wenig bis gar keinen monetären Wert hat.

Was die Aktualität der Daten betrifft, so sind die Informationen, die die Daten vermitteln, meist nur von begrenztem Wert. Beispielsweise sind Wetterdaten für heute oder morgen sehr relevant, aber sobald der entsprechende Tag vorbei ist, werden sie zu historischen Daten, die wesentlich weniger wertvoll sind. Darüber hinaus sind aktuelle Einkommensdaten auch deutlich wertvoller als historische Einkommensdaten, wenn es beispielsweise darum geht, Werbung gezielt auf die Konsumenten auszurichten. Allerdings können historische Daten je nach Geschäftsmodell auch einen signifikanten Wert haben, aber in den meisten Fällen wird die sogenannte Halbwertszeit als sehr kurz angesehen (Feijóo et al 2016).

Das bedeutet auch, dass sehr große Datenmengen nach kurzer Zeit wertlos sein können. Darüber hinaus gibt es bei den Unternehmen große Unterschiede in der Kompetenz, aus den Daten Erkenntnisse (relevante Informationen, Muster, Strukturen und

Trends) zu gewinnen (Banko & Brill 2001, Junqué de Fortuny et al 2013, Tucker 2010).¹³ Während diese Faktoren den Rahmen für die Diskussion um die Datenökonomie bilden sollten, werden sie bei anderen, mehr oder weniger nützlichen Datentypologien, tendenziell übergangen. Im folgenden Abschnitt beleuchten wir diese Datentypologien und ihren Einfluss auf die Diskussion.

2.4 Datentypologien

***Erkenntnisse:** Um Daten sinnvoll zu typologisieren, muß man ihren spezifischen Kontext und vorgesehenen Verwendungszweck mit einbeziehen. Im Allgemeinen ist es hilfreich, zwischen 1) strukturierten Daten, 2) unstrukturierten, sich wiederholenden Daten und 3) unstrukturierten, sich nicht wiederholenden Daten zu unterscheiden.*

Datentypologien sind ein entscheidender Aspekt des Datenaustauschs und für die Datenökonomie von wesentlicher Bedeutung, insbesondere bei der Festlegung potenzieller Einschränkungen für die Datenerfassung und -nutzung. Das offensichtlichste Beispiel für solche Einschränkungen ist der rechtliche Rahmen für die Erfassung und Nutzung personenbezogener Daten. Andere Typologien wie freiwillige, abgeleitete und beobachtete Daten, die kürzlich von Crémer et al (2019)¹⁴ vorgestellt wurden, können bei der Entwicklung eines Ordnungsrahmens zur Förderung einer erfolgreichen europäischen Datenökonomie eine Rolle spielen und sollten daher weiter untersucht werden. In dieser Studie konzentrieren wir uns auf eine Auswahl von Datentypologien, die für die aktuelle Diskussion über die sozioökonomischen Auswirkungen von Daten und ihren Einfluss auf den Wettbewerb im digitalen Zeitalter besonders relevant erscheinen.

Die am häufigsten verwendete Datentypologie stammt aus dem Kontext des Datenschutzes, insbesondere aus der kürzlich in Kraft getretenen allgemeinen EU-Datenschutzverordnung (GDPR).¹⁵ Hier wird zwischen personenbezogenen Daten und nicht personenbezogenen Daten unterschieden.¹⁶ Diese Dichotomie ist auch in vielen anderen Zusammenhängen von besonderer Relevanz. Insbesondere ist sie auch für die

¹³ Vgl. auch die Datenprodukte von Cao in Abschnitt 2.2.

¹⁴ Insbesondere Davis M, Martinez R, Kalaboukis C. 2010. Rethinking Personal Information – Workshop Pre-read, Invention Arts and World Economic Forum, Cologny führte erstmals diese Kategorisierung in Bezug auf die Art und Weise ein, wie Organisationen persönliche Daten erfassen können. World Economic Forum. 2011. Personal Data: The Emergence of a New Asset Class, WEF, Cologny. Die Kategorisierung war also nie dazu gedacht, verschiedene "Formen" von Daten zu beschreiben, noch war sie dazu gedacht, über den Bereich der persönlichen Daten hinaus verwendet zu werden.

¹⁵ Regulation (EU) 2016/679.

¹⁶ Personenbezogene Daten sind definiert als "alle Informationen, die sich auf eine identifizierte oder identifizierbare natürliche Person (im Folgenden „betroffene Person“) beziehen; als identifizierbar wird eine natürliche Person angesehen, die direkt oder indirekt, insbesondere mittels Zuordnung zu einer Kennung wie einem Namen, zu einer Kennnummer, zu Standortdaten, zu einer Online-Kennung oder zu einem oder mehreren besonderen Merkmalen, die Ausdruck der physischen, physiologischen, genetischen, psychischen, wirtschaftlichen, kulturellen oder sozialen Identität dieser natürlichen Person sind, identifiziert werden kann".

Verordnung über den freien Verkehr nicht-personenbezogener Daten¹⁷ von Bedeutung, die sich explizit auf nicht-personenbezogene Daten bezieht.¹⁸ Vor dem Hintergrund immer größer werdender Datenmengen und der Verbesserung der Möglichkeiten, verschiedene Datensätze zu kombinieren und zu analysieren, kann man bezweifeln, dass diese Dichotomie noch immer existiert. Insbesondere die Pseudonymisierung und Anonymisierung von Datensätzen, die offensichtlich personenbezogene Daten enthalten, scheint mehr denn je eine trügerische Vorstellung zu sein. Es herrscht eher ein Wettrennen zwischen Technologien, die eine Entpersonalisierung von Daten versprechen, und anderen, die diesen Prozess wieder rückgängig machen. Gegen Ende der 2000er Jahre war die Deanonymisierung oft möglich (Ohm 2010). Eine kürzlich durchgeführte Studie von Rocher et al (2019) zeigt, dass mit Hilfe von 15 demographischen Attributen 99,98% der Amerikaner in jedem Datensatz korrekt identifiziert werden könnten. Diese Ergebnisse lassen Zweifel daran aufkommen, ob selbst stark anonymisierte Datensätze modernen Anonymisierungsstandards, die z.B. die Datenschutzgrundverordnung vorschreibt, gerecht werden können. Damit stellen sie auch die technische und rechtliche Angemessenheit des De-Identifikations-Release-and-Forget-Modells in Frage.

Die Typologie der freiwilligen, abgeleiteten und beobachteten Daten, wie sie von Crémer et al (2019) vorgeschlagen wird, weist ebenfalls einige Grauzonen auf, wie die Autoren einräumen (S. 25). Diese Typologie eignet sich aus verschiedenen Gründen nicht für die Entwicklung eines regulatorischen Rahmens für die Datenökonomie. Wie von Hjørland (2019)¹⁹ veranschaulicht, kann dasselbe Datum in einem bestimmten Kontext der Datenerfassung freiwillig angegeben, aus anderen Daten abgeleitet oder aus Beobachtung in einem anderen Kontext gewonnen werden. Außerdem bezieht sich diese Typologie (wie ursprünglich von Davis et al (2010)²⁰ gemeint) ausschließlich auf den Prozess der Datenerfassung. In Anbetracht der Arbeiten von Drucker (2011) und Gitelman (2013) kann sogar argumentiert werden, dass alle digitalen Daten praktisch als abgeleitete Daten betrachtet werden sollten. Über diese grundsätzlichen Überlegungen zur Gültigkeit einer solchen Typologie hinaus kann es sinnvoll sein, den Pro-

¹⁷ Regulation (EU) 2018/1807.

¹⁸ Die Europäische Kommission definiert nicht-personenbezogene Daten im Zusammenhang mit der Verordnung über den freien Datenverkehr im Gegensatz zur Definition der personenbezogenen Daten im GDPR: *"Handelt es sich bei Daten um keine „personenbezogenen Daten“ im Sinne der Datenschutz-Grundverordnung, so sind sie als nicht-personenbezogen anzusehen. Nicht-personenbezogene Daten lassen sich je nach Herkunft klassifizieren als: Erstens: Daten, die sich ursprünglich nicht auf eine identifizierte oder identifizierbare natürliche Person beziehen, z.B. Daten über Wetterbedingungen, die von Sensoren generiert werden, die auf Windturbinen installiert sind, oder Daten über den Wartungsbedarf industrieller Maschinen; Zweitens: Daten, die ursprünglich personenbezogene Daten waren, später jedoch anonymisiert wurden. Die „Anonymisierung“ personenbezogener Daten unterscheidet sich von der Pseudonymisierung (siehe oben), denn ordnungsgemäß anonymisierte Daten können nicht einmal durch die Verwendung zusätzlicher Daten einer bestimmten Person zugeordnet werden und sind daher nicht-personenbezogene Daten"*. Siehe Europäische Kommission (2019): Leitlinien zur Verordnung über einen Rahmen für den freien Verkehr nicht-personenbezogener Daten in der Europäischen Union. COM(2019)250 final. S. 5–6.

¹⁹ Siehe Abschnitt 2.2.

²⁰ Vgl. World Economic Forum. 2011. Personal Data: The Emergence of a New Asset Class, WEF, Cologny; siehe auch Fußnote 7 oben.

zess der Datenerfassung im Hinblick auf die Datenqualität²¹ zu betrachten, die offensichtliche Auswirkungen auf den Wettbewerb hat.²²

Es gibt zahlreiche Ansätze zur Datentypisierung, die unter anderem die folgenden umfassen, aber nicht unbedingt darauf beschränkt sind (e.g. Kitchin 2014):

- quantitative (numerische, diskrete oder kontinuierliche) und qualitative (nichtnumerische) Daten.²³
- Nominal-, Ordinal- und Intervalldaten.
- Primär-, Sekundär- und Tertiärdaten.
- Erfasste, ausgeschöpfte, dynamische und abgeleitete Daten.
- Index-, Attribut- und Metadaten.
- unverarbeitete und verarbeitete Daten.
- relationale und multidimensionale Daten.
- gestreamte, gestapelte und gespeicherte Daten.
- streng geheime, hochsensible, sensible, private und öffentliche (unklassifizierte) Daten.

Eine sinnvolle Typologie der verschiedenen Datenarten hängt immer vom spezifischen Kontext und dem Verwendungszweck ab. Dennoch empfehlen wir, in Übereinstimmung mit dem Ansatz der Informationswissenschaft, Daten in erster Linie in Form von 1) strukturierten Daten, 2) unstrukturierten, sich wiederholenden Daten und 3) unstrukturierten, sich nicht wiederholenden Daten zu betrachten (Salinas & Lemus 2017). Tabelle 2–1 fasst einige wesentliche Merkmale dieser grundlegenden Datentypologie zusammen.

²¹ Dies wird im folgenden Abschnitt vertieft.

²² Diese Implikationen werden in Kapitel 4 vertieft diskutiert.

²³ Vor dem Hintergrund der Diskussion in Abschnitt 2.1 kann man hier „Dokumente“ als übergeordneten beschreibenden Begriff ansehen.

Table 2-1: Strukturierte und unstrukturierte Daten

Datentyp	Merkmale	Umfang	Verarbeitung	Speicherung	Nutzer
Strukturierte Daten (SD)	Diese Daten können durch vordefinierte Strukturen (Vektoren, Grafiken, Tabellen, u.a.) dargestellt werden. Die Struktur kann verallgemeinert werden.	Diese Daten gehören in die Domäne der traditionellen Datenbanksysteme und Data Warehouses.	Diese Daten können in Datenstrukturen wie Tabellen oder Arrays gespeichert und über weit verbreitete Sprachen wie SQL verwaltet werden.	Diese Daten werden in der Regel über relationale Datenbanken gespeichert und verwaltet.	Wirtschaftsanalysiker
Unstrukturierte sich wiederholende Daten (US-RD)	Diese Daten haben keine vordefinierte Struktur, sind zeitlich wiederkehrend, umfangreich. Nicht alle diese Daten haben einen Wert für die Analysen, daher können Proben oder Teile davon verwendet werden.	Diese Daten stammen von elektronischen Sensoren, deren Aufgabe die analoge Analyse des Signals ist, wie z.B. Vitaldaten, seismische Bewegungen, Positionierung, biologische und chemische Prozesse.	Im Allgemeinen gibt es definierte Algorithmen zur Behandlung dieser Art von Daten, wie z.B. die Fourier-Analyse für die Signale. Diese Art von Daten eignet sich für Wiederholungen und Wiederverwendung.	Diese Daten werden roh und kontextfrei gespeichert; dies geschieht mit Hilfe von NoSQL-Datenbanken (u.a. dokumentenorientiert, Key-Value, etc.) und Flat-Files.	Data-Mining-Experten für verschiedene Bereiche
Unstrukturierte, sich nicht wiederholende Daten (US URD)	Diese Daten haben keine einheitliche Struktur.	Sie umfassen Textinformationen, Bildanalyse, Dialoganalyse, Videoinhaltsanalyse und Stringanalyse.	Die Algorithmen für die Verarbeitung dieser Art von Daten sind nicht wiederverwendbar und die bloße Feststellung ihrer Struktur ist bereits eine komplexe Aufgabe. Je nach Art der Daten ist eine unterschiedliche Verarbeitung erforderlich, wie z.B. Verarbeitung natürlicher Sprache und computerlinguistische Techniken für textartige Daten.	Sie werden roh und kontextfrei in NoSQL-Datenbanken und Flatfiles gespeichert.	Experten für Datentechnik

Quelle: Salinas and Lemus (2017).

Während diese drei Datentypen bereits einige Informationen über die Qualität der betrachteten Daten implizieren (s. Table 2-1), gibt es viele weitere Aspekte, die im Hinblick auf die Datenqualität berücksichtigt werden müssen. Der folgende Abschnitt enthält eine detaillierte Darstellung dieser Aspekte und berücksichtigt mögliche Abwägungen zwischen Datenqualität und Datenmenge.

2.5 Datenmenge versus -qualität?

***Erkenntnisse:** Die richtige Balance zwischen Datenqualität und Datenmenge zu finden, ist eine ziemlich komplexe Aufgabe. Je nach Kontext und Zweck der Datennutzung müssen Unternehmen verschiedene Dimensionen der Datenqualität wie Genauigkeit, Vollständigkeit und Aktualität berücksichtigen, wobei eine mindestoptimale Datenmenge bei gleichzeitig abnehmendem Grenzertrag einer zunehmenden Datensatzgröße zu berücksichtigen ist. Daher müssen Unternehmen, die Big Data nutzen, verschiedene Abwägungen vornehmen.*

Die **Datenmenge** scheint im Mittelpunkt der öffentlichen Diskussion über Daten und ihre wirtschaftliche Nutzung zu stehen. So wird zumindest implizit eine Erhöhung der Datenmenge, auf die eine Organisation zugreifen kann – sei es durch die Erfassung von Daten, die eigene Erhebung von Daten oder die Erhebung durch Dritte – als positiv wahrgenommen. Häufig wird die Datenqualität ignoriert oder einfach als selbstverständlich angenommen. Dieser Abschnitt beleuchtet diese Annahmen und zeigt mögliche Wechselwirkungen zwischen Datenqualität und Datenmenge auf.

In der Praxis scheint die **Datenqualität** ein viel dringlicheres Problem zu sein als die Datenmenge. Laut Schroeder (2016) verbringen die meisten Data Scientists zwischen 75% und 90% ihrer Zeit mit der Bereinigung, Manipulation, Transformation und Vorbereitung von Daten für die Analyse, und dennoch wirkt sich eine schlechte Datenqualität nach wie vor nachteilig auf die Wirtschaft aus. IBM schätzt, dass der wirtschaftliche Schaden durch schlechte Daten in den USA jährlich 3,1 Billionen US-Dollar beträgt (Redman 2016). Schlechte Datenqualität hat enorme Auswirkungen auf Unternehmen. Beispielsweise können Entscheidungen, die auf schlechten Daten basieren, direkt zu Kundenunzufriedenheit, erhöhten Kosten und geringerer Arbeitszufriedenheit der Mitarbeiter führen, was sich letztlich auf die Gesamtleistung und den Umsatz des Unternehmens auswirken kann (Haug et al 2011, Redman 1998).

Im Gegensatz zu weit verbreiteten Vorstellungen können innovative Algorithmen aus dem Bereich der künstlichen Intelligenz die Mangelhaftigkeit der Daten nicht beheben. Wie Qi et al (2018) zeigen, erfordern unterschiedliche Anomalien in Daten unterschiedliche Anpassungen im Vorgehen, selbst für Standardaufgaben des maschinellen Lernens wie Klassifikation oder Clustering, die wiederum nur mit dem Wissen über den der jeweiligen Anomalie zugrundeliegenden Prozess möglich sind. Während viele Probleme mit fehlenden oder "verunreinigten" Daten zusammenhängen, können die Lösungen in ihrer Komplexität erheblich variieren, was sie mit der aktuell verfügbaren Rechenleistung undurchführbar machen kann. Die Experimente von Qi et al (2018) zeigen, dass neben den Daten selbst auch Metadaten, Domänenwissen und methodisches Fachwissen für Unternehmen, die die Datenanalyse unter Einbeziehung des maschinellen Lernens durchführen, gleichermaßen wichtig sind.

Es ist offensichtlich, dass Datenqualität vielschichtig ist (Pipino et al 2002, Sidi et al 2012, Taleb et al 2015, Taleb et al 2018, Wang & Strong 1996).²⁴ Wissenschaftler folgen hinsichtlich der verschiedenen Dimensionen der Datenqualität der grundlegenden Arbeit von Wang and Strong (1996). Diese sind in Tabelle 2–2 zusammengefasst und in Anlehnung an die Ergebnisse von Cichy and Rass (2019) angepasst, die die verschiedenen Datenqualitätsrahmen untersucht haben, die einzelne Dimensionen der Datenqualität berücksichtigen (siehe Abbildung. 2–3).

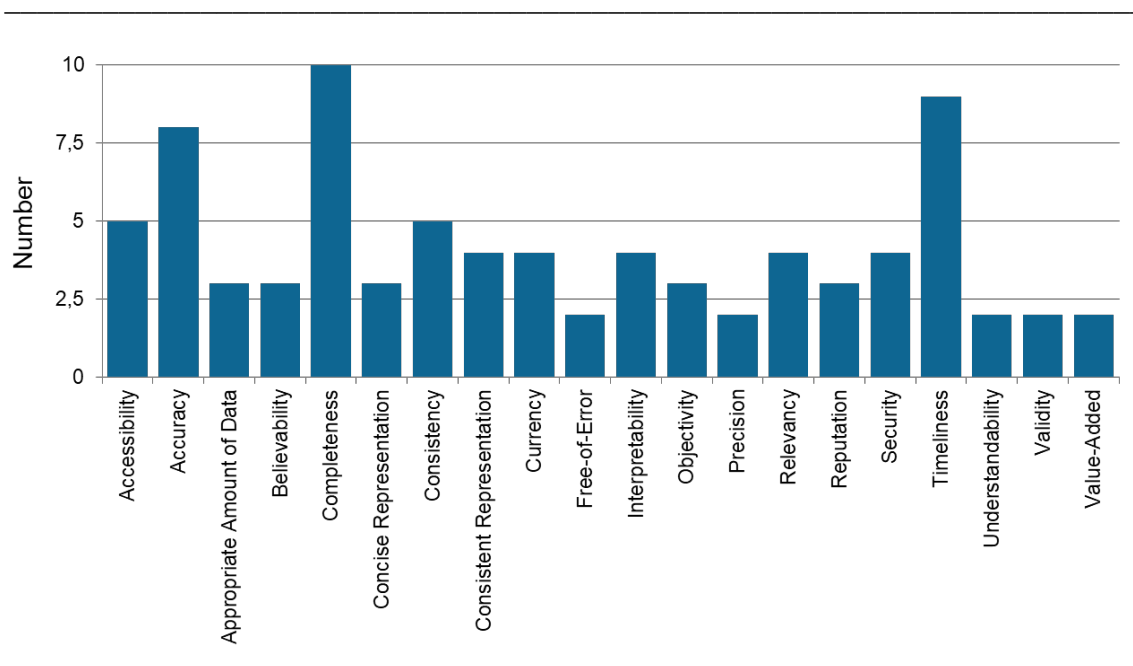
Tabelle 2-2: Wesentliche Dimensionen der Datenqualität

Zugriffssicherheit	Der Zugang zu den Daten kann eingeschränkt werden
Zugänglichkeit	Daten sind verfügbar oder leicht abrufbar
Genauigkeit	Daten sind korrekt und zuverlässig
Angemessene Datenmenge	Die Menge der verfügbaren Daten ist weder zu niedrig noch zu hoch
Glaubwürdigkeit	Daten werden als wahr und glaubwürdig angesehen
Vollständigkeit	Umfang und Detaillierungsgrad der Daten werden an die Aufgabe angepasst
Übersichtliche Darstellung	Daten werden kompakt und vollständig gespeichert
Leichte Verständlichkeit	Daten sind klar und leicht verständlich
Interpretierbarkeit	Daten sind klar definiert und in derselben Sprache und Einheit dargestellt
Objektivität	Daten sind objektiv und neutral
Relevanz	Daten können für eine bestimmte Aufgabe verwendet werden
Repräsentative Konsistenz	Daten sind im gleichen Format und kompatibel mit früheren Daten
Reputation	Die Datenquellen haben eine hohe Vertrauenswürdigkeit
Pünktlichkeit	Das Alter der Daten ist dem Zweck angepasst
Wertschöpfung	Daten liefern einen Mehrwert

Quelle: Nach Wang and Strong (1996).

²⁴ Insgesamt wurden in der wissenschaftlichen Literatur mehr als 170 für die Datenqualität relevante Dimensionen identifiziert. Ihre spezifische Relevanz hängt vom Kontext und dem Zweck der Datennutzung ab. Sidi F, Panahy PHS, Affendey LS, Jabar MA, Ibrahim H, Mustapha A. *2012 International Conference on Information Retrieval & Knowledge Management* 2012: 300-04. IEEE. unterstreichen diese Abhängigkeit in ihrem Überblick über die Dimensionen der Datenqualität, die direkt von der konkreten Aufgabe abhängig sind, die mit Hilfe der Daten gelöst werden soll. Solche Dimensionen könnten im weitesten Sinne unter dem Label "Eignung" subsumiert werden und umfassen z.B. Nützlichkeit, Abdeckung und Datenspezifikation.

Abbildung 2-3: Anzahl der Datenqualitätsrahmen und der entsprechenden Dimensionen der Datenqualität



Quelle: Cichy and Rass (2019).

Taleb et al (2015) und Taleb et al (2018) nutzten diese Erkenntnisse, um ein ganzheitliches Datenqualitätsmanagement-Modell für die Big Data-Wertschöpfungskette zu entwickeln. Es umfasst acht Stufen von der Datenerfassung bis zur Visualisierung. Auf jeder dieser Stufen kann die Datenqualität beeinträchtigt werden. Es ist auch offensichtlich, dass die Stufen voneinander abhängig sind und die Datenqualität nur so gut sein kann wie das schwächste Glied in dieser Kette. Dies bedeutet, dass ein Unternehmen, das nur wenig über den spezifischen Datenerfassungsprozess weiß, bei der Nutzung der Daten wahrscheinlich vor Herausforderungen steht. Dasselbe gilt, wenn die Datenübertragung unzuverlässig ist, z.B. bei hoher Paketverlustrate. Lenart et al (2018) weisen darauf hin, dass selbst ein einzelner Sensor je nach (Teil-)Aufgabe unterschiedlich leistungsfähig sein kann. Innerhalb eines Netzwerks aus zahlreichen Sensoren wird es noch schwieriger, die Glaubwürdigkeit eines bestimmten Datums festzustellen. Während konsistente Ergebnisse über verschiedene Sensoren hinweg die Glaubwürdigkeit der Daten stärken, wird sie durch widersprüchliche Informationen geschwächt. Die Herausforderung besteht dann darin, zu entscheiden, welcher der Sensoren nicht richtig funktioniert und warum.²⁵

Je nach Kontext und Zweck der Datenauswertung kann die **Menge der Daten** es Organisationen ermöglichen, Daten, die nicht den für die Aufgabe festgelegten Qualitätskriterien entsprechen, zu übergehen oder zu manipulieren. Mehr Daten gehen in der Re-

²⁵ In der Forschung beziehen sich diese Fragen auf Konzepte der internen und externen Reliabilität sowie teilweise auf die interne und externe Validität von Daten.

gel mit der Bereitstellung von mehr Informationen einher, so dass die Lernkurve eines datenbasierten Geschäftsmodells im Vergleich zu traditionellen Geschäftsmodellen exponentiell ansteigt. (Junqué de Fortuny et al 2013). Dies hängt allerdings vom Kontext der Datenerfassung und -nutzung ab.

Die Entwicklung des Konzepts von Big Data von einem rein quantitativen zu einem heute allgemein akzeptierten, viel weiter gefassten Konzept spricht jedoch dafür, dass die Quantität als solche stark mit den qualitativen Merkmalen der Daten verflochten ist. Während sich das vorherrschende Verständnis von Big Data eine Zeit lang direkt (und ausschließlich) auf die Menge der Daten bezog, die erfasst und analysiert oder anderweitig verwendet werden, gibt es einen Konsens darüber, dass die Merkmale von Big Data durchaus von der Menge bestimmt werden können, aber sicherlich viele andere Faktoren umfassen.

Um den Blick auf Big Data zu erweitern, liefern de Mauro et al (2016) eine Definition, die die Komplexität und den Umfang von Big Data zusammenfasst: *"Big Data sind der Informationswert, der sich durch eine so große Menge, Geschwindigkeit und Vielfalt auszeichnet, dass für seine Umwandlung in Wertschöpfung spezifische Technologien und Analysemethoden erforderlich sind."*²⁶ Sie definieren Big Data nicht nur durch die Datenmenge, sondern berücksichtigen auch die verschiedenen Arten von verfügbaren Daten und die Geschwindigkeit, mit der die Daten erfasst und analysiert werden. Darüber hinaus berücksichtigt die Definition die Bedeutung der Technologien und Methoden, die notwendig sind, um diese Daten verwertbar und damit wertvoll zu machen.²⁷ Big Data stellt jedoch besondere Herausforderungen an die Datenqualität. Insbesondere angesichts der großen Menge der gewonnenen Daten, der hohen Geschwindigkeit des Dateneingangs und der großen Datenvielfalt wird von einer unzureichenden Datenqualität ausgegangen.

Tatsächlich ist eines der größten Probleme im Zusammenhang mit Big Data das Informationsrauschen. Diesbezüglich kommen Liu et al (2016) zu folgendem Schluss: *"Obwohl Big Data Informationen in [...] detaillierter Form enthalten, speichert es auch zufällige Schwankungen, Fluktuationen und sogar Geräusche während der Messung. Bei der Anwendung [...] des maschinellen Lernens zur Analyse von Big Data stoßen Wissenschaftler oft auf das Phänomen des Over-Fittings, bei dem der Algorithmus des maschinellen Lernens aus dem in den feinkörnigen Daten eingebetteten Störgeräusch lernt und auf der Grundlage der gestörten Information eine Vorhersage trifft"* (S. 138). Eine der schwerwiegendsten Folgen ist, dass möglicherweise fehlerhafte Muster oder Wechselwirkungen festgestellt werden. Auch ohne Verzerrungen im Datensatz können aufgrund der schiereren Datenmenge Korrelationen zwischen einzelnen Variablen im Daten-

²⁶ Es gibt Studien, die weitere Eigenschaften wie Wert und Wahrhaftigkeit erfassen (siehe z.B. White).

²⁷ Für eine Definition mit den typischen Vs von Big Data siehe z.B. Gandomi A, Haider M. 2015. Beyond the hype: Big data concepts, methods, and analytics. *International journal of information management* 35: 137-44. Insbesondere Boyd D, Crawford K. 2012. Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon. *Information, Communication & Society* 15: 662-79 bieten eine kritischere Darstellung von Big Data als die gängigen Branchennarrative. Sie wurden in Abschnitt 2.2 zitiert und werden deshalb hier nicht wiederholt.

satz gefunden werden, die theoretisch nicht korrelieren sollten (Boyd & Crawford 2012, Fan et al 2014). Liu et al (2016) weisen auf ein weiteres Problem hin, nämlich die Herkunft der Daten. Daten, die z.B. von kommerziellen Datenanbietern erworben wurden, sind oft nicht brauchbar. Diese neigen dazu, bei der Auswahl und Erhebung von Daten keine wissenschaftlichen Methoden zu verwenden. In den meisten Fällen handelt es sich bei den erfassten Personengruppen um eine kleine Anzahl von Personen, die eine bestimmte Dienstleistung oder ein bestimmtes Gerät nutzen (eine verzerrte Stichprobe). Studien, die dies nicht berücksichtigen, können damit zu Annahmen gelangen, die nicht für die Gesamtbevölkerung gelten. Dies gilt insbesondere für Daten, die aus Social Media stammen (Blank & Lutz 2017). Eine weitere zentrale Herausforderung im Zusammenhang mit solchen Daten besteht darin, dass sie keine Qualitätsbezüge aufweisen. Dies legt nahe, dass *"Daten in der Anfangsphase profiliert und mit bestimmten Qualitätsinformationen versehen werden müssen. Dies bedeutet auch, dass die Qualität der Datenattribute während ihres gesamten Lebenszyklus bewertet, verbessert und kontrolliert werden muss, da sie sich direkt auf die Ergebnisse der Analysephase auswirkt"* (Chen et al 2014).

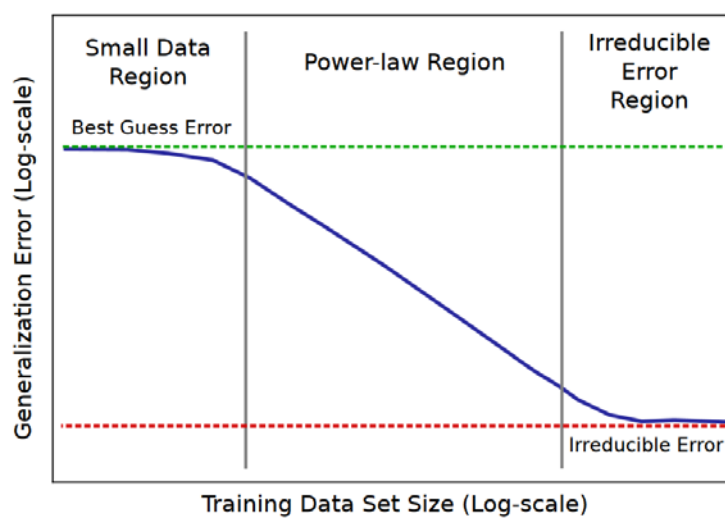
Unabhängig von der Quantität und Qualität ist zu beachten, dass *"große Datenmodelle auf Korrelation und nicht auf Kausalität basieren, so dass sie nicht außerhalb des Datenbereichs extrapoliert werden können. Außerdem ist es schwierig, Datenfehler, Unsicherheit und Messrauschen von den tatsächlichen Phänomenen zu trennen. Anlagendaten können verwendet werden und wurden auch schon verwendet, um First-Principles-Modelle zu optimieren. Ein richtig abgestimmtes First-Principles-(Causation)-Modell ist immer besser als ein Korrelationsmodell, das aus der Analyse großer Datenmengen entwickelt wurde"* (Saudagar et al 2019).

In Bezug auf die erforderliche Datenmenge weisen Pipino et al (2002) darauf hin, dass diese weder zu gering noch zu groß sein sollte. Sie vermeiden es jedoch, eine spezifische Datenmenge anzugeben, da sich die ausreichende Menge an Dateneinheiten zwangsläufig von einem Kontext und Zweck zum anderen unterscheiden wird. Es ist nicht unbedingt von vornherein erkennbar, wieviel Daten zur Lösung einer bestimmten Aufgabe oder eines Problems genau benötigt werden. Auch kann die Menge heute, da die Digitalisierung den Zugriff auf potenziell extrem große Datensätze ermöglicht, weniger wichtig sein als zu Beginn der 2000er Jahre. Die öffentliche Diskussion konzentriert sich heute auf die für bestimmte Aufgaben erforderliche Mindestdatenmenge und auf die Frage, ob es bei den Daten abnehmende Skalenerträge gibt. Während die verbreitete Vorstellung davon ausgeht, dass mehr Daten immer besser sind, weist die Forschung auf einen abnehmenden Nutzen bei zunehmender Größe der Datensätze hin (Li et al 2016).

Kürzlich haben Hestness et al (2017) die Auswirkungen von zusätzlichen Daten auf vier gängige Anwendungen des Deep Learning verglichen: (1) Maschinelle Übersetzung, (2) Sprachmodellierung, (3) Bildverarbeitung und (4) Spracherkennung. Dabei entdeckten sie vergleichbare Muster im Hinblick auf die Verbesserungen der Modellgenauigkeit

(vgl. Abbildung 2-4). Ihre Ergebnisse deuten darauf hin, dass es tatsächlich eine Mindestgröße für die Menge an Trainingsdaten gibt, mit der Deep-Learning-Modelle in den Power-Law-Bereich der Lernkurve gelangen. Wenn die Trainingsdatensätze sehr groß werden, verflacht sich die Lernkurve. Die Autoren beschreiben diesen Bereich als den nicht reduzierbaren Fehlerbereich. Diese Untergrenze hat ihren Ursprung in statistischen (Bayes-)Fehlern und auch in Fragen der Datenqualität, wie z.B. falsch gekennzeichnete Proben in den Trainings- oder Validierungsdaten. Der genaue Exponent der Power-Law-Region hängt stark vom Kontext und Zweck der Lernaufgabe sowie von der Datenqualität ab. Schließlich weisen die Autoren darauf hin, dass größere Trainingsdatensätze einen großen Teil der Genauigkeit kompensieren können, die durch weniger genaue Berechnungen möglicherweise verlorengelht. Die Datenmenge kann jedoch keine grundlegenden Mängel in der Datenqualität ausgleichen, wie es der nicht reduzierbare Fehlerbereich suggeriert. Über diese Überlegungen hinaus äußern Schwartz et al (2019) Bedenken, dass die ständig steigenden Trainingsdaten den CO₂-Fußabdruck der Künstlichen Intelligenz (KI) unangemessen erhöhen und anspruchsvolle Anwendungen der KI im Grunde genommen zu einer Aktivität einer ausgewählten Gruppe von Elite-Universitäten und großen Unternehmen machen. Vor diesem Hintergrund sollten Wissenschaftler, die sich mit der Entwicklung von Algorithmen beschäftigen, die Größe der verwendeten Datensätze optimieren.

Abbildung 2-4: Skizze der Power-Law-Lernkurven



Quelle: Hestness et al (2017).

Die wichtigsten Ergebnisse aus Kapitel 3

- *Daten durchdringen praktisch alle Wirtschaftsaktivitäten.*
- *Die Datenökonomie ist kein Sektor, der sich trennscharf von anderen (traditionellen) Sektoren abgrenzen lässt.*
- *Die Wertschöpfung erfolgt in Zyklen und Netzwerken und nicht in sequenziellen Interaktionen entlang einer Wertschöpfungskette.*
- *Der Prozess der Wertschöpfung, der mit Daten beginnt, kann vollständig innerhalb eines einzigen Unternehmens stattfinden oder viele Akteure einbeziehen.*
- *Es ist abzuwägen zwischen sinkenden Kosten für die Datenerfassung und steigenden Kosten für die effektive Kontrolle über die Daten.*
- *Unternehmen müssen erhebliche organisatorische Maßnahmen ergreifen, um die Möglichkeiten der Datennutzung (Big Data) voll auszuschöpfen.*
- *Big Data kann Mustererkennung, Vorhersageanalysen und neue Erkenntnisse ermöglichen.*
- *Daten an sich machen eine digitale Plattform nicht erfolgreich. Es gibt andere Faktoren, die eine ebenso wichtige Rolle spielen.*
- *Die Rolle von Daten bei der Innovation ist eine dreifache: (1) Daten können ein Treiber für Produkt- und Prozessinnovation sein; (2) Daten sind ein integraler Bestandteil wissensbasierter Produktinnovationen; und (3) kontinuierliche Daten aus der Praxis können neue Möglichkeiten zur Messung der Innovationsaktivität und des Erfolgs in der amtlichen Statistik bieten.*
- *Die größte Herausforderung bei der Nutzung von Daten bei Innovationen ist in der Regel nicht der unmittelbare Zugang zu den Daten, sondern die Überwindung der technologischen und prozessbedingten Herausforderungen bei der Datennutzung.*
- *Die Rolle von Daten bei der Innovation ist bei "datenarmen" und "datenreichen" Kontexten unterschiedlich.*

3 Wertschöpfung aus Daten

Die Wertschöpfung aus Daten bildet logischerweise das Kernziel der Datenökonomie. In diesem Kapitel beleuchten wir zunächst die grundsätzliche Funktionsweise der Datenökonomie, die sich aufgrund der Fließfähigkeit der Daten von den traditionellen Konzepten der ökonomischen Wertschöpfung unterscheidet. Im Anschluss untersuchen wir die konkrete wirtschaftliche Datennutzung und die unterschiedlichen Herausforderungen, die erfüllt sein müssen, um ihre Möglichkeiten voll auszuschöpfen. Schließlich untersuchen wir die Rolle von Daten für Innovation.

3.1 Datenökonomie

***Erkenntnisse:** Die Datenökonomie besteht aufgrund der Fluidität der Daten aus zahlreichen komplexen Interaktionen zwischen den Marktakteuren. Als solche passt sie nicht in traditionelle Wertschöpfungskettenkonzepte, sondern folgt einer zirkulären Logik, bei der verschiedene Kreisläufe der Wertschöpfung zu Wertschöpfungsnetzwerken verflochten sind.*

Wie im vorhergehenden Kapitel erläutert, können aus den erfassten und/oder erhobenen Daten oder einer Kombination davon durch die Generierung von Erkenntnissen mittels Datenanalyse wirtschaftlich verwertbare Informationen gewonnen werden. Um diesen Prozess der Wertschöpfung aus Daten zu ermöglichen, gibt es eine Reihe von Dienste- und Infrastrukturanbietern, die die Prozesse der Datenerfassung, -übertragung, -speicherung, -analyse und -verwertung unterstützen.²⁸

Dieser Prozess findet in allen Unternehmen aller Branchen auf unterschiedlichem Niveau statt. Die Datennutzung überschreitet somit die traditionellen Branchengrenzen, da mit zunehmender Digitalisierung die Kompetenzen, die früher (hauptsächlich) im Bereich der Informations- und Kommunikationstechnologie (IKT) vorhanden waren, praktisch alle Branchen durchdringen.²⁹ Folglich dürften enge Definitionen der Datenökonomie, wie

²⁸ Für weitere Beschreibungen dieser wirtschaftlichen Aktivitäten und ihrer Wechselwirkungen siehe z.B. BVDW. 2018. Datenwertschöpfung und Qualität von Daten, Bundesverband Digitale Wirtschaft (BVDW) e.V., Düsseldorf, GSMA. 2018. The Data Value Chain, GSMA, Curry E. 2016. The Big Data Value Chain: Definition, Concepts, and Theoretical Approaches In *New Horizons for a Data-Driven Economy - A Roadmap for Usage and Exploitation of Big Data in Europe*, ed. JM Cavanillas, E Curry, W Wahlster, pp. 29-38: SpringerOpen, Attard J, Orlandi F, Auer S. *International Conference on Theory and Practice of Electronic Governance, New Delhi, India, 2017*: 475-784..

²⁹ Wie Calvino F, Criscuolo C, Marcolin L, Squicciarini M. 2018. A taxonomy of digital intensive sectors, Organisation for Economic Co-operation and Development, Paris zeigen, unterscheiden sich die Branchen hinsichtlich ihrer Investitionen in digitale Technologien und Daten und deren voraussichtlicher Nutzungsmöglichkeiten. Ähnliche Ergebnisse finden sich in Arnold R, Schiffer M, Pols A. 2013. *Wirtschaft Digitalisiert - Welche Rolle spielt das Internet für die deutsche Industrie und Dienstleister?*, IW Consult and BITKOM, Cologne, Berlin.

sie vom DataLandscape-Projekt ³⁰ vorgeschlagen und von BEREC (2019)³¹ übernommen wurden, hinter der Realität zurückbleiben.

So wie die traditionellen Branchengrenzen die Datenökonomie nicht abbilden können, so ist auch das traditionelle Konzept der Wertschöpfungskette nicht auf die Datenökonomie anwendbar. Wie in Abschnitt 2.3 beschrieben, haben Daten viele ökonomische Eigenschaften, die nicht mit den traditionellen Konzepten von Exklusivität, Konsum oder Produktwert übereinstimmen. Vor allem sind Daten nicht-rival in der Nutzung und nicht-ausschließbar, so dass die Datenökonomie nicht als typische Wertschöpfungskette verstanden werden sollte, an deren Ende ein Produkt gekauft und konsumiert oder verwendet wird.³² In Übereinstimmung mit Arnold and Waldburger (2015) sollte die Datenökonomie als ein Kreislauf der Wertschöpfung betrachtet werden (siehe Abbildung 3–1).

In ihrem Konzept, das als "Daten-Wertschöpfungskreislauf" bezeichnet wird, sind die Nutzer (d.h. Verbraucher und/oder Unternehmen) sowohl der Endpunkt als auch der Ausgangspunkt der Wertschöpfung. Einerseits sind sie die Informationsquellen, die eine Wertschöpfung im Verlauf des Kreises ermöglichen. Zum anderen sind sie letztlich die Empfänger von Dienstleistungen und Produkten, die auf diesen Daten basieren. Dazwischen gibt es verschiedene Stufen der Wertschöpfung. Zum einen gibt es digitale Geräte, die Daten erfassen.³³ Dann werden die Daten über Kommunikationsnetze übertragen. Die Stufe des Datenhandlings umfasst alle Dienste, die sich mit der Speicherung, Analyse und sonstigen Verarbeitung von Daten befassen. Innerhalb der Diensteebene werden die durch die Datenanalyse gewonnenen Informationen genutzt, um bestehende Prozesse, Produkte und Dienstleistungen zu verbessern, effektiver zu gestalten oder neue zu entwickeln.

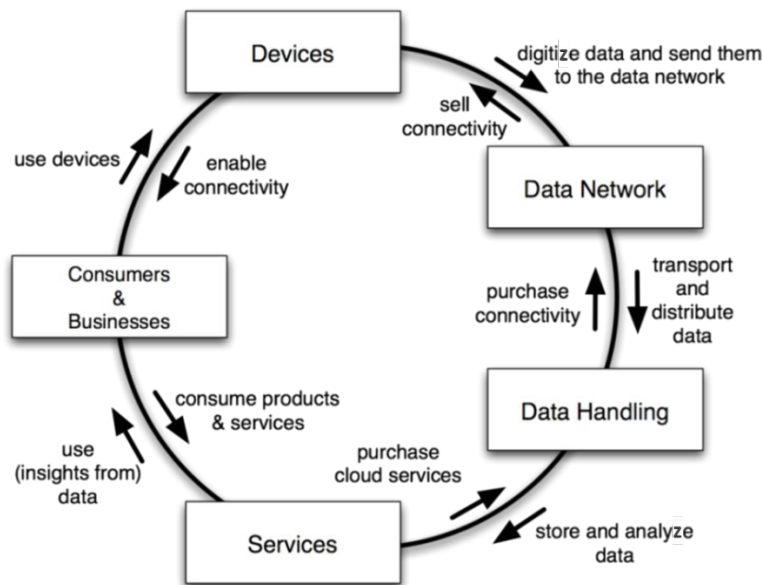
³⁰ <http://datalandscape.eu/> Die Website sowie die dort präsentierten Berichte und Daten sind Teil der EU-Datenmarktstudie, die von der Europäischen Kommission im Rahmen von SMART 2013/0063 in Auftrag gegeben wurde.

³¹ Der Begriff "Data Economy" umfasst die (wachsende) Verfügbarkeit von Daten, die damit verbundenen unternehmerischen Möglichkeiten sowie den (potenziellen) gesellschaftlichen Wert der generierbaren Erkenntnisse. Laut dem Bericht der Europäischen Kommission "Building a European Data Economy" werden mit der "Data Economy die gesamten Auswirkungen des Datenmarktes – d.h. des Marktes, auf dem digitale Daten als Produkte oder als aus Rohdaten erzeugte Dienstleistungen ausgetauscht werden – auf die Wirtschaft als Ganzes erfasst. Sie umfasst die Erzeugung, Sammlung, Speicherung, Verarbeitung, Verteilung, Analyse, Aufbereitung, Bereitstellung und Nutzung von Daten mit Hilfe digitaler Technologien. Eine wichtige Entwicklung in der Data Economy der letzten Jahre ist die Zunahme der Vielfalt und des Volumens der Daten, die durch Online-Aktivitäten erzeugt werden" (S. 7). Siehe auch: European Commission. 2017. Communication from the Commission to the European Economic and Social Committee and the Committee of the Regions - "Building a European Data Economy" (SWD(2017) 2 final), European Commission, Brussels.

³² Das traditionelle Konzept der Wertschöpfungskette wird üblicherweise zurückgeführt auf Porter ME. 1985. *Competitive Advantage: Creating and sustaining superior performance*. New York: The Free Press.

³³ Wie bereits in Abschnitt 2.1 dargelegt, kann die Datenerfassung auch durch Software erfolgen. Während die Geräte die ultimative Schnittstelle der Dateneingabe bleiben, führt die Software im Hintergrund die notwendigen Prozesse zur Digitalisierung, Vorverarbeitung und Speicherung der Daten durch. Daher kann der Zugriff auf die Software ausreichen, um Zugang zu den von einem Gerät erfassten Daten zu erhalten.

Figure 3-1: Daten-Wertschöpfungskreislauf



Quelle: Arnold and Waldburger (2015).

Der Daten-Wertschöpfungskreislauf veranschaulicht die zentralen Komponenten und Wechselwirkungen der verschiedenen Organisationen bzw. Instanzen in den fünf Phasen. Alle Stufen des Daten-Wertschöpfungskreislaufs können jedoch Teil ein und desselben Unternehmens sein. Offensichtlich sind große Online-Plattformen in verschiedene Stufen des Daten-Wertschöpfungskreislaufs eingetreten, da sich ihre Reichweite mittlerweile von Geräten (Smartphones, Tablets, Set-Top-Boxen usw.) über Datennetzwerke, einschließlich Unterwasserkabel, bis hin zur Datenverarbeitung mit hochentwickelten Cloud-Infrastrukturen und Diensten für Verbraucher und Unternehmen erstreckt.

Große Unternehmen aus traditionellen Industriezweigen haben sich ähnlich verhalten. Sie haben auch verschiedene Stufen des Daten-Wertschöpfungskreislaufs in ihre Geschäftsmodelle integriert. So haben beispielsweise Originalausrüstungshersteller (OEMs) aus dem Automobilsektor die Konnektivität in ihre Fahrzeuge integriert, eigene Rechenzentren und Datenanalyse-Tools aufgebaut und einen zunehmenden Anteil ihrer Einnahmen aus Dienstleistungen generiert. Wir beobachten ähnliche Trends bei Landwirtschaft, Maschinen und Chemikalien.³⁴

³⁴ Der steigende Anteil der dienstleistungsbezogenen Wertschöpfung in den traditionellen Industriesektoren ist seit der zweiten Hälfte der 2000er Jahre zu beobachten. Daten, *Digitization* und Digitalisierung haben diesen Prozess ermöglicht. Erste Ausführungen zu diesem Trend finden sich in Kempermann H, Lichtblau K. 2012. Definition und Messung von hybrider Wertschöpfung. *IW Trends* 39: 1-20 und Lichtblau K, Arnold R. 2012. Smart Industry – Intelligente Industrie: Eine neue Betrachtungsweise der Industrie. Ergebnisse einer Studie der Institut der deutschen Wirtschaft Köln Consult GmbH für das Land Hessen, Initiative Industriepark Hessen, Neu-Isenburg.

Ungeachtet der von den Geräten erfassten Datenart, beleuchtet der Daten-Wertschöpfungskreislauf das von Carrière-Swallow and Haksar (2019) verwendete Konzept der Datengerechtigkeit.³⁵ Organisationen, die Daten für ihre eigenen Zwecke erfassen oder verwerten, dürften ein Interesse daran haben, den Zugang zu diesen Daten zu kontrollieren. Ihre Kontrolle kann dadurch eingeschränkt sein, dass auch andere Organisationen dieselben (oder vergleichbare) Geräte zur Datenerfassung verwenden können. Da ein heutiges Gerät vermutlich eine Kombination von Teilen verschiedener Lieferanten und ebenso eine Kombination von Software verschiedener Anbieter ist, ist es auch wahrscheinlich, dass jede dieser Organisationen in der Lage und willens sein könnte, einige der Daten, die das Gerät erfasst, zu erfassen und zu nutzen.

Die Erfassung und Integration der Daten verschiedener Geräte über mehrere Lieferketten hinweg steht im Mittelpunkt der Vision von Industrie 4.0. Diese Vision verspricht zwar beträchtliche wirtschaftliche Vorteile, doch die abnehmende Kontrolle der Unternehmen über ihre Daten und das mangelnde Vertrauen unter den Wettbewerbern können den Erfolg dieser Vision letztendlich zunichte machen. Letztendlich hängt der Erfolg von Industrie 4.0 davon ab, ob die Kosten der Datenerfassung und der wirtschaftliche Wert der generierten Informationen die Kosten ausgleichen können, die durch die zunehmende Komplexität der Kontrolle über die Daten für die beteiligten Unternehmen entstehen.³⁶

Wertschöpfungsnetzwerke sind ein weiterer zentraler Bestandteil der Vision von Industrie 4.0. Sie werden durch die Fluidität der Daten ermöglicht und können als Kombination verschiedener (Teilbereiche des) Daten-Wertschöpfungskreislauf(s) betrachtet werden. Attard et al (2016, 2017) veranschaulichen dies in ihrem Konzept der Wertschöpfung aus Daten: Das sogenannte Daten-Wertschöpfungsnetzwerk (DVN).³⁷ Ein DVN wird definiert als *"eine Reihe voneinander unabhängiger Maßnahmen mit dem Ziel, aus Daten Wert zu schaffen, indem sie als Produkt verwertet werden, an dem verschiedene Akteure [...] beteiligt sein können, die eine oder mehrere Aktivitäten durchführen [...] und jede Aktivität kann aus einer Reihe von Aktionen oder wertschöpfenden Maßnahmen bestehen [...]. Jede Aktion kann wiederum aus einer oder mehreren Daten-Wertschöpfungsketten bestehen, da sie unter Umständen eine bestimmte Abfolge von Prozessen erfordern, die nacheinander ausgeführt werden müssen"* (Attard et al 2017).

³⁵ Die Perspektive von Carrière-Swallow Y, Haksar V. 2019. The Economics and Implications of Data - An Integrated Perspective, Washington, DC bleibt auf personenbezogene Daten beschränkt, bei denen es um die betroffene Person und die mögliche Weiterverwendung der personenbezogenen Daten durch den Datensammler und den Datenverarbeiter geht.

³⁶ Es ist denkbar, dass die Kosten für einen wirksamen Schutz des Datenzugriffs letztendlich den potentiellen Ertrag, der mit der Erfassung der Daten erzielt werden kann, übersteigen. Es besteht also auch die Möglichkeit, dass das in der Data Economy beobachtete Wachstum nicht unbedingt eine Einbahnstraße ist.

³⁷ Merkmale des DVN sind: nicht-materielles Datenprodukt, nicht-sequentiell, mehrere Akteure, verschachtelte Wertschöpfungsketten, wiederkehrendes Wertschöpfungsnetzwerk, unabhängige Aktivitäten.

Im Wesentlichen haben Daten als solche wenig oder keinen Wert. Stattdessen können sie in wirtschaftlich verwertbare Informationen umgewandelt werden, die dann im Rahmen von Geschäftsmodellen genutzt werden. Im folgenden Abschnitt werden solche datengesteuerten Geschäftsmodelle näher betrachtet.

3.2 Geschäftsmodelle in der Datenökonomie

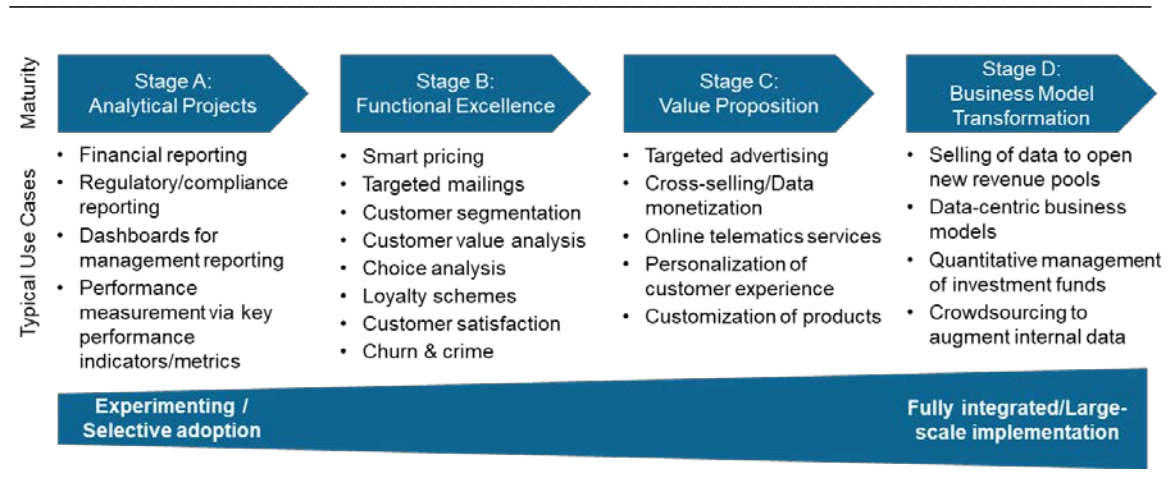
Erkenntnisse: *Die Nutzung von Daten für geschäftliche Zwecke erfordert einen erheblichen Aufwand auf Seiten des Unternehmens, der weit über den Datenzugriff hinausgeht. Um die Daten in vollem Umfang nutzen zu können, müssen die Unternehmen eine erhebliche Transformation bis hin zu einer vollständigen Umgestaltung ihres Geschäftsmodells vollziehen.*

Die Digitalisierung eröffnet den Unternehmen neue Möglichkeiten zur Optimierung ihres Geschäftsgebarens, und der Zugang zu Daten ist dabei von zentraler Bedeutung, wobei die Vorreiter in der Digitalisierung tendenziell führend in der Verwendung von Daten sind. Eine Umfrage der OECD (2015) zur IKT-Nutzung zeigt einige betriebswirtschaftliche Argumente und potenzielle Auswirkungen der Datennutzung in Unternehmen auf. Eine Organisation kann die Möglichkeiten der Datenanalyse nutzen, um (1) potenzielle Kunden zu identifizieren, (2) die Kaufbereitschaft der Kunden durch gezielte Angebote und Rabatte zu erhöhen, (3) Produkte (Waren und Dienstleistungen) auf die Bedürfnisse der Kunden zuzuschneiden und (4) die Effektivität der betrieblichen Produktion und/oder Organisation zu steigern. Die (positiven) Auswirkungen der Datenanalyse für Unternehmen sind daher wahrscheinlich eine Mischung aus (1) potenziellen Kosteneinsparungen, (2) Umsatzwachstum und (3) Verbesserungen in der Unternehmensorganisation.

In ähnlicher Weise identifizieren Grover et al (2018) vier Hauptziele bei der Nutzung von Big Data in Unternehmen. Erstens kann Big Data durch die Verbesserung der unternehmerischen Entscheidungsfindung einen Mehrwert schaffen. Dies kann erreicht werden, indem ein breiter und konsistenter Datenzugang im gesamten Unternehmen gewährleistet wird, ergänzt durch Entscheidungsstrukturen, um auf der Grundlage der Daten handeln zu können, oder durch Entscheidungsmodelle, die die menschliche Entscheidungsfindung unterstützen oder in Geschäftsprozesse integriert sind. Zweitens kann Big Data einen Mehrwert schaffen, indem die Effektivität, Effizienz und Produktivität von Geschäftsprozessen verbessert werden, was zu einer Verbesserung der Abläufe und zu einem geringeren Zeitaufwand für Prozessausfälle führt. Drittens kann Big Data einen Mehrwert bei der Produkt- und/oder Dienstleistungsinnovation schaffen. Viertens kann Big Data auch zu einem besseren Einkaufserlebnis und wettbewerbsfähigeren Dienstleistungen führen, was wiederum eine höhere Kundenzufriedenheit und -bindung zur Folge hat. Das von Tiefenbacher and Olbrich (2015) entwickelte Reifegradmodell für Big Data geht sogar noch weiter, da es (neue oder erweiterte) Mehrwer-

te und sogar eine vollständige Geschäftsmodelltransformation als Stufen im Rahmen der Nutzung von Big Data in Unternehmen empfiehlt (s. Abbildung 3–2).

Abbildung 3-2: Reifegradmodell für Big Data



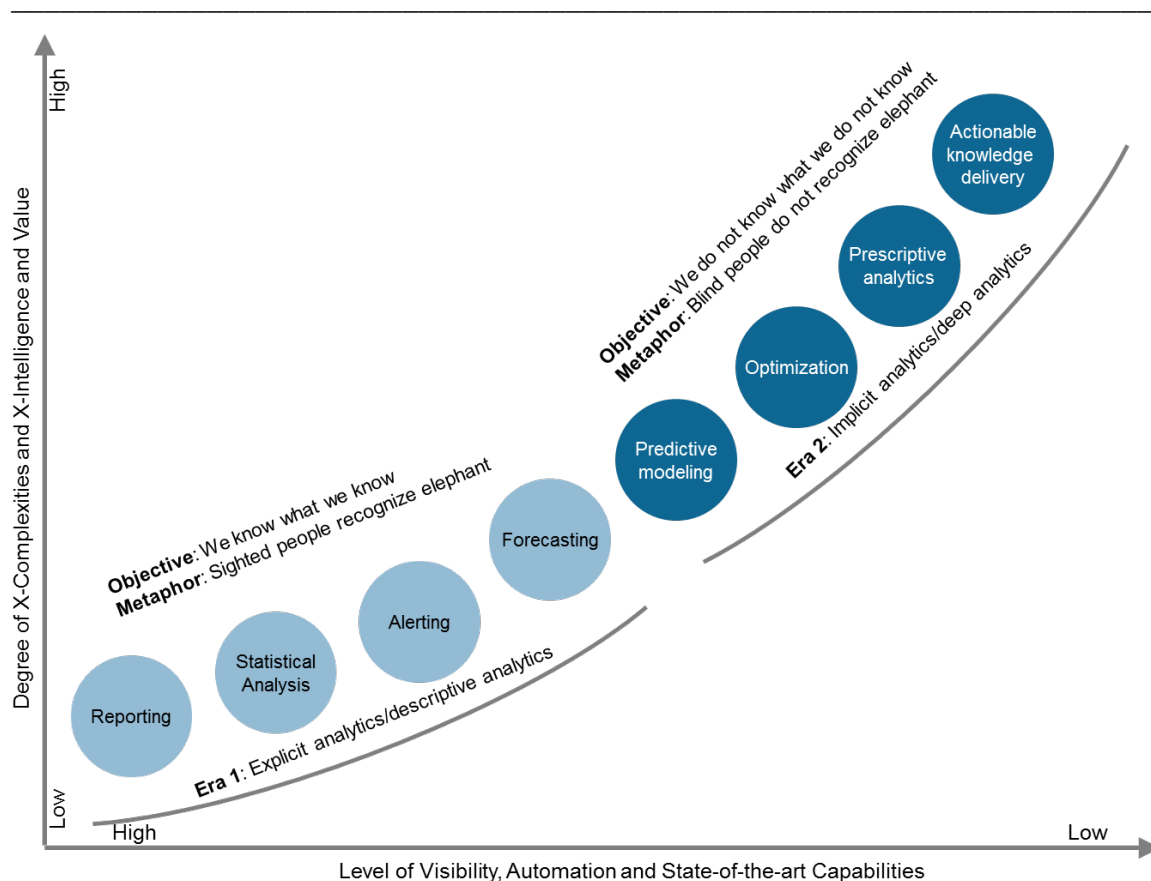
Quelle: Tiefenbacher and Olbrich (2015).

In Hinblick auf die Wertschöpfung aus Daten (Big Data) gibt es zwei sozio-technische Faktoren, wie Unternehmen diese realisieren, nämlich Portabilität und Interkonnektivität (Günther et al., 2017). Portabilität bezieht sich auf die Fähigkeit, auf digitale Daten aus einem Anwendungskontext zuzugreifen und diese in einen anderen Kontext zu übertragen. So können Daten über Plattformen und Unternehmensgrenzen hinweg übertragen und zugänglich gemacht werden. Interkonnektivität bezieht sich auf die Fähigkeit, Daten aus verschiedenen Quellen zu kombinieren und damit Entscheidern und Experten durch die Erschließung von Verknüpfungen, Mustern und Zusammenhängen weitere Erkenntnisse zu ermöglichen. Auf diese Weise können durch die Nutzung von Big Data neue Mehrwerte erzielt werden.

Cao (2017) konzentriert sich besonders auf den spezifischen Mehrwert, den die Datenanalyse in Unternehmen schaffen kann. Hierbei unterscheidet er zwei Phasen. Zum einen die Phase der expliziten und weitgehend deskriptiven Analytik, die (noch) den Datenanalysefähigkeiten der meisten Unternehmen entspricht. Die Phase der impliziten und detaillierten Analyse hingegen erweitert die Möglichkeiten der Datenanalyse über die Beschreibung von bereits Bekanntem hinaus auf das Wie und Warum von Phänomenen der realen Welt. Nach Cao (2017) können die Erkenntnisse aus der impliziten und detaillierten Analyse dazu genutzt werden, die nächstbeste oder schlechteste Situation im Vergleich zum aktuellen Zustand zu ermitteln, um optimale Interventionsstrategien innerhalb oder zwischen den beteiligten Parteien zu entwickeln. Im Hinblick auf die in den Abschnitten 2.2 und 2.5 vorgestellte Diskussion erscheinen diese Erwartungen etwas überzogen, da sie zumindest implizit von einer theorieunabhängigen Generierung von Erkenntnissen aus Daten ausgehen, die nicht möglich ist. Dennoch verdeutlicht der

in Abbildung 3-3 dargestellte allgemeine Evolutionspfad der Datenanalyse einen wichtigen Punkt in Bezug auf deren Transparenz, indem die Transparenz mit zunehmender Komplexität abnimmt, ein Ergebnis, das etwas widersprüchlich ist. Mit anderen Worten, je ausgefeilter und umfassender die Datenanalyse ist, desto undurchsichtiger wird die eigentliche Erkenntnisgenerierung. Die Algorithmen werden immer komplexer und der Automatisierungsgrad der Datenanalyse steigt mit dem Einsatz von Werkzeugen wie z.B. dem Deep Learning. Ein linearer Entwicklungspfad der Datenanalyse innerhalb einer Organisation ist unwahrscheinlich, da laut Cao (2017) viele Hürden auf dem Weg zu Rückwärts- und Vorwärtsiterationen führen, einschließlich einer beträchtlichen Anzahl von Trial-and-Error-Zyklen innerhalb der Wissenschaftler-Teams.

Abbildung 3-3: Spektrum und Entwicklung der Datenanalytik von explizit bis implizit



Quelle: Cao (2017).

Natürlich kommen all diese (positiven) Ergebnisse nur dann zum Tragen, wenn die Organisation in der Lage ist, den Informationswert der von ihr durch Datenanalyse und -erhebung erfassten Daten gezielt zu erschließen. Laut McKinsey (2016) sind komplexe organisatorische Veränderungen erforderlich, um Daten (Big Data) innerhalb eines Unternehmens voll nutzbar zu machen. Im Einzelnen beschreiben die Autoren fünf Stufen: (1) Entwicklung von Anwendungsfällen; (2) Aufbau eines Daten-Ökosystems; (3) Er-

werb von analytischen Fähigkeiten, die zur Ableitung von Erkenntnissen aus Daten erforderlich sind; (4) Änderung von Geschäftsprozessen, um Erkenntnisse aus Daten in den aktuellen Arbeitsablauf zu integrieren; (5) Entwicklung der Kompetenzen von Führungskräften und Beschäftigten im mittleren Management, damit diese verstehen, wie datengestützte Erkenntnisse innerhalb des aktuellen Geschäftsmodells oder bei der Entwicklung eines neuen Geschäftsmodells genutzt werden können. Um die Vorteile von Daten (Big Data) in vollem Umfang nutzen zu können, kann laut Günther et al. (2017) das Erreichen von funktionaler Kompetenz innerhalb der Organisation nur der erste Schritt zur Transformation des Geschäftsmodells sein.

Die Umwandlung von (Teilen) des Unternehmens in eine digitale Plattform kann ein mögliches Ziel der oben skizzierten Geschäftsmodelltransformationen sein. Da digitale Plattformen auch einen Schwerpunkt der aktuellen öffentlichen Debatte über Datenzugang und -nutzung darstellen, werden sie hier näher untersucht, wobei wir uns auf die Rolle der Daten innerhalb digitaler Plattformen konzentrieren.

Diese Rolle hängt entscheidend von dem spezifischen Konzept der digitalen Plattformen ab. Drei Grundkonzepte digitaler Plattformen werden typischerweise diskutiert: (1) digitale Plattformen als wirtschaftliches Konzept, (2) digitale Plattformen als technologische Plattform und (3) digitale Plattformen als Vermittler und Anbieter von (Medien-) Inhalten. Aufgrund ihrer kreisförmig angeordneten Stufen passt die Datenökonomie besonders gut zum ersten dieser Konzepte.³⁸

Gemäß der wirtschaftswissenschaftlichen Literatur über Plattformen als wirtschaftliches Konzept (Armstrong 2006, Caillaud & Jullien 2003, Evans & Noel 2005, Evans & Schmalensee 2007, Hagiu 2007, Rochet & Tirole 2003, Rysman 2009) haben diese die folgenden Merkmale gemeinsam:

- (1) Digitale Plattformen fungieren als Vermittler und ermöglichen die Interaktion verschiedener Nutzergruppen. Die Dienste der digitalen Plattformen (Vermittlung, Transaktion, Austausch, Vergleich) orientieren sich am Verhalten und den Nutzungsmustern der jeweiligen Nutzergruppe. Eine direkte Beteiligung der Plattform an der Interaktion ist nicht zwingend erforderlich. Die Erfassung und Nutzung von Daten ist jedoch eindeutig notwendig, um diese Schlüsselfunktion einer digitalen Plattform zu erfüllen.
- (2) Diese Interaktionen sind durch sogenannte direkte und indirekte Netzwerkeffekte gekennzeichnet. Wie in Abschnitt 2.3 dargelegt, vergrößern diese Effekte wahrscheinlich die Chance des Unternehmens, eine zunehmende Menge und möglicherweise bessere Qualität von Daten zu sammeln, die wiederum zur weiteren Verbesserung der im Rahmen der digitalen Platt-

38 Dies wird weiter ausgeführt in Arnold R, Waldburger M. 2015. The Economic Influence of Data and their Impact on Business Models In *Trends in Telecommunication Reform 2015 - Getting Ready for the Digital Economy*, ed. ITU, pp. 153-83. Geneva: International Telecommunication Union.

form angebotenen Dienste genutzt werden können, insbesondere im Hinblick auf die unter (1) angesprochenen spezifischen Zielgruppen.

- (3) Die Preisgestaltung von digitalen Plattformen ist in der Regel eine Funktion der Preiselastizitäten der Nachfrage verschiedener Plattform-Nutzergruppen. Die Elastizitäten spiegeln auch die gegenseitigen Abhängigkeiten der einzelnen Seiten der Plattform wider. Jene mit geringer Preiselastizität neigen dazu, einen relativ hohen Preis zu zahlen und damit überwiegend die Kosten der jeweiligen Plattform zu decken, während jene mit hoher Preissensibilität dazu neigen, einen niedrigen oder gar keinen Preis für die Plattformnutzung zu zahlen (d.h. Quersubventionierung). Die Vergütung kann in Form von Geld, Daten und Aufmerksamkeit für Werbung erfolgen.

Digitale Plattformen als technologische Plattformen werden im Wesentlichen über ihre Funktionalitäten definiert. Die Plattform stellt eine einheitliche technische Basis dar, auf der Komponenten verbunden und (Software-)Programme betrieben werden können. Generell kann zwischen einer Hardware-Plattform und einer Software-Plattform unterschieden werden: Eine Hardware-Plattform (auch Maschinenebene genannt) besteht aus einer Rechnerarchitektur und den zugrunde liegenden Befehls- und Prozessorstrukturen. Im Gegensatz dazu bildet eine Software-Plattform (auch Anwendungsebene genannt) die Grundlage, auf der Betriebssysteme und Anwendungsprogramme betrieben werden können. In diesem Sinne spielt eine digitale Plattform eine entscheidende Rolle, um die Datenerfassung auf standardisierte Weise zu ermöglichen. Solche technologischen digitalen Plattformen und die von ihnen erfassten Daten können ein integraler Baustein eines Geschäftsmodells sein, insbesondere wenn sie den Zugang zu Daten ermöglichen, aus denen wirtschaftlich verwertbare Informationen generiert werden können.

Das Konzept der digitalen Plattformen als Aggregatoren von Inhalten basiert auf dem Plattformkonzept der Medienindustrie. Ihre Kernfunktionalität besteht in der Bereitstellung digitaler Übertragungskapazitäten oder digitaler Datenströme (auch von Dritten) mit dem Ziel, diese Angebote den Nutzern des Dienstes als Gesamtpaket zugänglich zu machen. Sie entscheiden auch über die Auswahl für die Aggregation. Die Qualität der Plattform wird also im Wesentlichen dadurch bestimmt, wie ein Unternehmen über die Zusammenstellung der Inhalte entscheidet. Daten sind daher ein integraler Bestandteil des Geschäftsmodells solcher Plattformen.³⁹

Bei allen drei Arten von digitalen Plattformen zeigt sich, dass Daten zwar in jedem Fall eine Rolle spielen, diese Rolle aber je nach Kontext sehr unterschiedlich ist. Darüber hinaus ist es offensichtlich, dass neben dem Datenzugriff auch andere Faktoren der Geschäftsmodellgestaltung über Erfolg oder Misserfolg eines Unternehmens entschei-

³⁹ Wir erläutern dies in Abschnitt 3.3.

den.⁴⁰ Innovation ist ein Schlüsselfaktor für den langfristigen Erfolg und steht im Mittelpunkt des folgenden Abschnitts.

3.3 Innovation in der Datenökonomie

Erkenntnisse: Die Rolle von Daten bei der Innovation ist eine dreifache: (1) Daten können eine treibende Kraft für Produkt- und Prozessinnovationen sein; (2) Daten sind ein integraler Bestandteil wissensbasierter Produktinnovationen; und (3) kontinuierlich erhobene Daten aus der Praxis können neue Ansätze zur Messung von Innovationsaktivitäten und -erfolg in der amtlichen Statistik liefern. Die größte Herausforderung bei der Nutzung von Daten für Innovationen ist in der Regel nicht der Zugang zu Daten durch eigene Datenerhebung oder die Datenerhebung durch Dritte, sondern die Überwindung der technologischen und prozessbedingten Herausforderungen für ihre Nutzung.

Innovationsforscher sind sich einig, dass die Innovationstätigkeit und letztlich erfolgreiche Innovationen von komplexen Interaktionen zwischen privaten und öffentlichen Akteuren abhängen; dies wird als "Innovationssystem" betrachtet.⁴¹ Diese inhärente Komplexität macht es fast unmöglich, spezifische Triebkräfte für erfolgreiche Innovationen zu identifizieren. Kausalzusammenhänge zwischen einem bestimmten Input und Output sind noch schwerer zu erkennen.

Dennoch ist es offensichtlich, dass Daten für praktisch jede Innovationstätigkeit wesentlich waren und bleiben werden. Folglich weisen die OECD und Eurostat in ihrer jüngsten Ausgabe des Oslo-Handbuchs⁴² (OECD & Eurostat 2018) darauf hin, dass es für die Innovationsstatistik keinen Mehrwert darstellt, zu beobachten, ob Daten an einer bestimmten Innovation beteiligt waren oder nicht. Stattdessen empfehlen die OECD und Eurostat, die allgemeine digitale Kompetenz jedes Unternehmens zu messen und diese mit seiner Innovationsaktivität und seinem Erfolg in Beziehung zu setzen. Diese Empfehlung deutet auf eine wichtige Einsicht für die gegenwärtige Debatte über den

⁴⁰ Wir erläutern dies näher in Arnold R, Bott J, Hildebrandt C, Schäfer S, Tenbrock S. 2016. Internetbasierte Plattformen und ihre Bedeutung in Deutschland, Wissenschaftliches Institut für Infrastruktur und Kommunikationsdienste (WIK), Bad Honnef und Arnold R, Hildebrandt C. 2017. The Socio-Economic Impact of Online Platforms, Wissenschaftliches Institut für Infrastruktur und Kommunikationsdienste (WIK), Bad Honnef.

⁴¹ "Innovationssysteme ist keine Wirtschaftstheorie im gleichen Sinne wie die neoklassische oder evolutionäre Ökonomie, sondern das Konzept integriert theoretische Perspektiven und empirische Erkenntnisse, die auf mehreren Jahrzehnten Forschung beruhen. Innerhalb dieses Ansatzes ist Innovation sowohl ein kumulativer, pfad- und kontextabhängiger Prozess als auch ein interaktiver Prozess", siehe Lundvall B-Å, Borrás S. 2005. Science, technology and innovation policy In *The Oxford handbook of innovation*, ed. J Fagerberg, DC Mowery, RR Nelson, pp. 599-631. Oxford: Oxford University Press cf. Bertenrath R, Arnold R, Koppel O, Lang T. 2011. Innovation Policy and the Business Cycle: Innovation Policy's Role in Addressing Economic Downturn - INNO-Grips Policy Brief No. 1, European Commission, Cologne/Brussels.

⁴² Das Oslo-Handbuch ist die wichtigste Informationsquelle dafür, wie Innovation in der amtlichen Statistik verstanden und gemessen wird.

Datenzugang und -austausch hin. Daten sind ein integraler Bestandteil jeder Innovationsaktivität, können aber nur dann einen Mehrwert schaffen, wenn das Unternehmen ausreichend in der Lage ist, sie zu nutzen. Folglich ist der Datenzugang eine notwendige, aber nicht hinreichende Bedingung für die Förderung der Innovationstätigkeit und letztlich des Wettbewerbs um die besten Produkte und Dienstleistungen.

Neben weiteren Faktoren der digitalen Leistungsfähigkeit empfiehlt das Oslo-Handbuch die Verwendung von Indikatoren für die Leistungsfähigkeit eines Unternehmens in Bezug auf *"den Zugang zu und die Fähigkeit zur Nutzung von Datenanalysen zur Gestaltung, Entwicklung, Vermarktung und Verbesserung von Produkten, einschließlich der Daten über die Nutzer der Unternehmensprodukte und deren Umgang mit diesen Produkten"*. (S. 123) Zumindest der "Zugang zu" Daten hängt entscheidend vom spezifischen Umfeld ab, in dem das Unternehmen tätig ist, da die **Datenfülle** in den verschiedenen Bereichen stark variiert (siehe Abschnitt 2.5). Insbesondere *"die Art der benötigten Daten unterscheidet sich von einer Branche zur anderen und oft auch je nach branchenspezifischer Nutzung. Auch Verfügbarkeit und Datenzugang sowie Datenqualität und die Möglichkeit der Integration verschiedener Datenbestände sind unterschiedlich"* (Paunov & Planes-Satorra 2019). Grundsätzlich können wir zwischen datenintensiven und datenarmen Kontexten unterscheiden.

In **datenintensiven Bereichen** wie z.B. Online-Diensten kann ein Unternehmen für seine Innovationsaktivitäten in der Regel auf eine Fülle von Daten zugreifen, die im Rahmen der weitgehend digitalen Interaktionen mit den Nutzern seiner eigenen Dienste (kontinuierlich) erfasst wurden. Mit Hilfe von Cookies, Geräte- oder Browser-Fingerabdrücken können Unternehmen sogar Erkenntnisse über das Verbraucherverhalten bei anderen Online-Diensten als den eigenen gewinnen.⁴³ Im Hinblick auf die Innovationstätigkeit im Zusammenhang mit Online-Diensten ist die A/B-Erprobung von (neuen oder verbesserten) Merkmalen in (Pseudo-)Experimenten eine maßgebliche Auswirkung der Digitalisierung und der Datennutzung. Da A/B-Tests und andere datengestützte Innovationsaktivitäten in datenintensiven Kontexten zu vernachlässigbaren Kosten durchgeführt werden können, sind alternative Innovationsaktivitäten sowohl kostenintensiver als auch weniger effektiv. Ein Unternehmen, das in einem datenintensiven Kontext konkurrenzfähig sein möchte, muss Daten zu einem zentralen Bestandteil seines Innovationsprozesses machen. Dies ist weder kostspielig noch schwierig. Allerdings wird entscheidend sein, dass das Unternehmen die von ihm benötigten Daten gezielt erfassen kann. Fraglich ist, ob ein solches Unternehmen einen wesentlichen

⁴³ Es gibt reichlich Literatur über die verschiedenen Möglichkeiten, Verbraucher online zu beobachten. Für einen Überblick siehe Boerman SC, Kruikemeier S, Zuiderveen Borgesius FJ. 2017. Online Behavioral Advertising: A Literature Review and Research Agenda. *Journal of Advertising* 46: 363-76. Für eine Perspektive, wie diese Ansätze in Geschäftsmodelle integriert werden, siehe Hildebrandt C, Arnold R. 2016. Big Data und OTT-Geschäftsmodelle sowie daraus resultierende Wettbewerbsprobleme und Herausforderungen bei Datenschutz und Verbraucherschutz - WIK-Diskussionsbeitrag Nr. 414, Wissenschaftliches Institut für Infrastruktur und Kommunikationsdienste (WIK), Bad Honnef oder Bott J, Hildebrandt C, Arnold R. 2018. Die Nutzung von Daten durch OTT-Dienste zur Abschöpfung von Aufmerksamkeit und Zahlungsbereitschaft: Implikationen für Wettbewerb, Regulierung sowie Daten- und Verbraucherschutz - WIK-Diskussionsbeitrag Nr. 431, Wissenschaftliches Institut für Infrastruktur und Kommunikationsdienste (WIK), Bad Honnef.

Nutzen aus dem Zugriff auf die von einem anderen (konkurrierenden) Unternehmen erfassten Daten zu ziehen.⁴⁴ Folglich müssen Unternehmen in datenintensiven Bereichen für ihre Innovationsaktivitäten selbst Daten erfassen, da sie die volle Kontrolle über die Quelle und Herkunft der Daten und die Geräte (oder besser gesagt die Messinstrumente) für die Datenerfassung haben müssen.⁴⁵ Darüber hinaus scheint der (Mehr)Wert des Zugriffs auf Daten Dritter begrenzt zu sein.

In **datenarmen Bereichen** wie Logistik und Transport oder Abfall und Recycling ist der Datenzugriff exponentiell komplexer und kostspieliger als in datenintensiven Bereichen, da viele Interaktionen zwischen den Akteuren immer noch offline stattfinden und nur wenige Prozesse digitalisiert werden. Folglich kann in datenarmen Bereichen die Realisierung des Datenzugriffs an sich als eine Innovationstätigkeit eines Unternehmens betrachtet werden. Während die Kosten für den Datenzugriff in solchen Bereichen höher sind, können die Erträge einer Innovation, die einen neuartigen Datenzugriff im Markt ermöglicht, ebenfalls wesentlich höher sein. Tatsächlich können solche Innovationen für einen traditionell datenarmen Sektor disruptiv sein. Folglich hat ein Unternehmen mit einer Innovationstätigkeit, die neue Datenzugriffsmöglichkeiten beinhaltet, wahrscheinlich einen starken Anreiz, seine Daten zu schützen, da diese wahrscheinlich der eigentliche Wettbewerbsvorteil der Innovation sind.

Innovation in datenintensiven Bereichen

In datenintensiven Kontexten sind Innovationen, die von Daten angetrieben werden, eher inkrementell. Dies ist vor allem darauf zurückzuführen, dass Unternehmen leicht auf große Datenmengen zugreifen können, was wiederum die Kosten für eine ständige und innovative Anpassung an selbst kleinste Marktveränderungen reduziert. Darüber hinaus haben gerade in der wissensbasierten, entmaterialisierten digitalen Wirtschaft selbst kleinere Innovationen einen entscheidenden Einfluss auf den unternehmerischen Erfolg und die Wettbewerbsfähigkeit. Eine der zentralen Methoden, die Unternehmen in datenintensiven Bereichen zur Innovation einsetzen, ist das A/B-Testing.

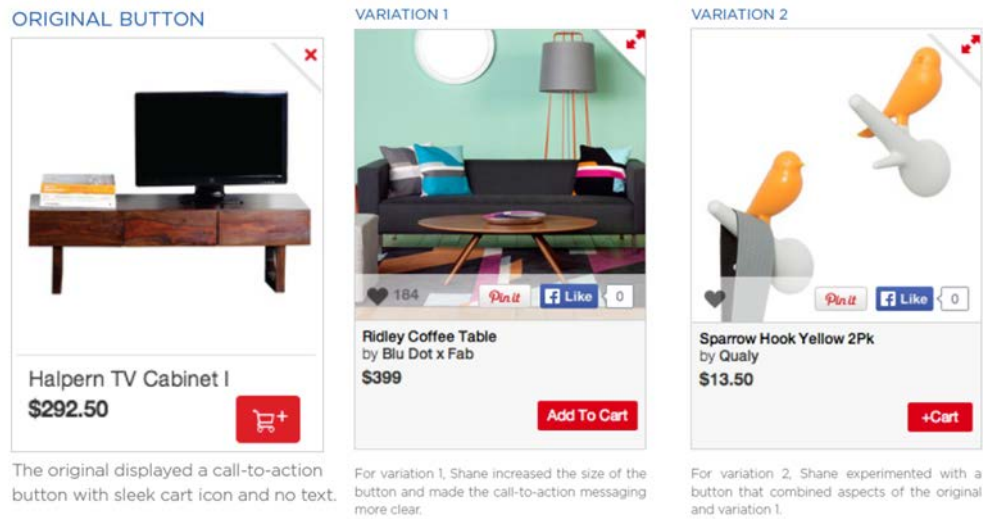
A/B-Tests werden von Website-Betreibern häufig verwendet, um das Design ihrer Website kontinuierlich zu verbessern und so das Nutzerinteresse zu erhöhen. Fab (<https://fab.com/>) hat zum Beispiel A/B-Tests verwendet, um verschiedene Arten von "In den Warenkorb"-Buttons zu testen. Fab ist eine E-Commerce-Plattform sowohl für den Verkauf als auch für den Kauf von Kleidung, Haushaltswaren, Accessoires und mehr. Die Tests umfassten drei Arten von Schaltflächen, und zwar eine auf Symbolen basierende Schaltfläche und zwei, die hauptsächlich textbasiert waren (

⁴⁴ Der A/B-Test ist ein gutes Beispiel dafür. Offensichtlich kann ein Unternehmen auf der Grundlage der über seine Nutzer erfassten Daten wertvolle Erkenntnisse für die Verbesserung seines eigenen Angebots gewinnen. Der Wert der Erkenntnisse, die aus Daten über die Nutzer anderer Dienste gewonnen werden, dürfte für den gleichen Zweck wahrscheinlich geringer oder sogar gleich Null sein.

⁴⁵ Es handelt sich eher um eine Software als um ein Hardware-Gerät.

Abbildung 3-4).

Abbildung 3-4: Beispiel für A/B-Tests



Die Änderung der Schaltfläche hatte einen nicht zu vernachlässigenden Einfluss auf die Klickraten. Variante 1 erhöhte die Leistung um 49%.⁴⁶ AutoScout24, ein Online-Marktplatz für Neu- und Gebrauchtwagen, wandte ebenfalls A/B-Tests an, um die Konversionsrate des Formulars zu erhöhen. Auch hier wurden mehrere Designs gegen das Original getestet. Im Ergebnis waren die neuen Varianten in Bezug auf die Umsetzungsrate dem Original überlegen.⁴⁷

⁴⁶ Eine umfassende Beschreibung des Verfahrens ist verfügbar unter <https://s3.amazonaws.com/optimizely-marketing/customer-stories/fab-casestudy.pdf> [Letzter Zugriff am 22.10.2019].

⁴⁷ Eine umfassende Beschreibung des Verfahrens ist verfügbar unter <https://www.optimizely.com/de/customers/autoscout24/> [Letzter Zugriff am 22.10.2019].

Innovation in datenarmen Bereichen

Ein Sektor, der immer noch als datenarm eingestuft werden kann, ist die Logistikbranche. Besonders innovative Unternehmen in diesem Sektor haben jedoch erkannt, welchen Wert digitale Daten haben können. Diese Firmen konzentrieren ihre innovative Tätigkeit auf die Erstellung digitaler Informationen, die ehemals analoge Prozesse mit überwiegend manueller Eingabe abbilden. Ein Beispiel für eine solche Firma ist **Carrypicker** (<https://www.carrypicker.com/>). Carrypicker ist ein Start-up-Unternehmen, das die Effizienz von Speditionsunternehmen steigern will. Ein großes Problem in diesem Bereich ist, dass der gesamte Planungsprozess trotz seiner hohen Komplexität selbst bei den größten Speditionen noch immer manuell durchgeführt wird. Um eine Umsatzmaximierung zu erreichen und freie Kapazitäten zu reduzieren, werden die neuesten Methoden aus den Bereichen Künstliche Intelligenz, maschinelles Lernen und Predictive Analytics eingesetzt. Ziel ist die Entwicklung einer dynamischen Online-Preisgestaltungsplattform sowie die Optimierung und Zuweisung von Routen.⁴⁸

Cirplus (<https://www.cirplus.io/>) ist ein weiteres Start-up, das den Datenzugriff in einem datenarmen Sektor verbessern will. Sein Schwerpunkt liegt auf der Recyclingindustrie und es entwickelt derzeit einen B2B-Marktplatz für recycelten Kunststoff. Nicht nur ist die Recyclingrate für Kunststoffverpackungen extrem niedrig, sondern die meisten Einwegprodukte bestehen auch aus verschiedenen Kunststoffmaterialien, was das Recycling und die Herstellung hochwertiger Recyclate erschwert. Gleichzeitig steckt die Digitalisierung dieser Industrie noch in den Kinderschuhen, was den Prozess sowohl für Käufer als auch für Verkäufer kompliziert macht. Der unabhängige Software-Dienstleister hat es sich daher zur Aufgabe gemacht, die gesamte Kunststoff- und Entsorgungsindustrie zu vernetzen – und dabei nicht nur eine Plattform zu bieten, sondern auch einen neuen Standard für die Kennzeichnung und den Informationsaustausch über Recyclate einzuführen.⁴⁹

Quelle: WIK-Consult.

Nachdem wir die grundlegenden Unterschiede in der Rolle, die Daten je nach der Datenfülle des jeweiligen Innovationsumfeldes spielen, herausgearbeitet haben, betrachten wir nun die spezifischen Auswirkungen von Daten auf die Innovation. Innovationen werden in erster Linie durch die zunehmende Digitalisierung ermöglicht. Entsprechend dem zu Beginn dieses Kapitels vorgestellten Daten-Wertschöpfungskreislauf weisen Paunov and Planes-Satorra (2019) darauf hin, dass *"intelligente und vernetzte Geräte eine ergiebige Quelle für Innovationen in allen Sektoren sind. Sie sammeln und übertragen Daten über Prozesse, Nutzung und Umweltbedingungen, die eine Prozessopti-*

⁴⁸ Eine umfassende Beschreibung des Verfahrens ist verfügbar unter <https://www.bmvi.de/SharedDocs/DE/Artikel/DG/mfund-projekte/carrypicker.html> [Letzter Zugriff am 22.10.2019].

⁴⁹ Für weitere Informationen siehe <https://www.capital.de/wirtschaft-politik/dieses-start-up-koennte-die-recyclingindustrie-revolutionieren> [Letzter Zugriff am 22.10.2019], <https://www.cirplus.io/pilot-program> [Letzter Zugriff am 22.10.2019].

rierung, vorausschauende Analytik/Diagnostik und in ihrem fortgeschrittensten Stadium den autonomen Betrieb von Systemen ermöglichen, wie dies bei selbstfahrenden Autos der Fall wäre" (S. 11). Darüber hinaus können diese Daten zusätzliche Innovationen ermöglichen, da sie neue Möglichkeiten bieten, Produkte und Dienstleistungen zu diversifizieren oder das Leistungsversprechen der Anbieter insgesamt zu verändern, d.h. sie von Produzenten zu Dienstleistern zu machen.⁵⁰ Und schließlich können Daten durch ihre Fluidität aus ehemals geschlossenen Sektorsilos "heraussickern". Dies ermöglicht es neuen Marktteilnehmern, einen bedeutenden Teil der Wertschöpfung in einem Sektor zu erzielen, sofern sie nur in der Lage sind, auf Daten zuzugreifen und diese auf innovativere Weise zu nutzen als die etablierten Dienstleistungen, Produkte oder Prozesse in diesem Sektor.

Typisches Beispiel – Uber

Uber ist ein offensichtliches Beispiel für ein Unternehmen, das bestehende Infrastrukturen und Technologien nutzt, um einen seit Jahrzehnten bestehenden Dienst auf innovative Weise anzubieten. Uber nutzte die bestehende Infrastruktur von Geräten wie Smartphones und die Datenübertragung über Telekommunikationsnetze, um eine Schnittstelle für den Datenaustausch zwischen Fahrern und Fahrgästen zu schaffen. Diese Schnittstelle ermöglicht die digitale Abwicklung der Leistungserbringung.

Über den kontinuierlichen Datenfluss von Geräten hinaus hat Uber auch auf bestehende und weitgehend statische Daten, wie z.B. Karten, zurückgegriffen. Diese Kombination führte zu reduzierten Warte- und Fahrzeiten. Als das Unternehmen zu wachsen begann, ersetzte oder ergänzte Uber öffentliche Daten durch eigene Daten, die es aufgrund erheblicher Investitionen in neue Technologien erfassen konnte.

Über diese drei grundlegenden Auswirkungen digitaler Daten auf die Innovation hinaus können ihre Auswirkungen bei verschiedenen Arten von Innovationen, wie sie im Oslo-Handbuch (OECD & Eurostat 2018) definiert sind, beobachtet werden: Grundsätzlich lassen sich Unternehmensinnovationen in **Produktinnovationen** und **Prozessinnovationen** untergliedern. Diese werden weiter in Untertypen eingeteilt, die in den folgenden Abschnitten diskutiert werden, wobei der Schwerpunkt auf den Auswirkungen liegt, die digitale Daten auf jeden dieser Typen haben.

⁵⁰ Dies wird auch als hybride Wertschöpfung bezeichnet. Siehe Kempermann H, Lichtblau K. 2012. Definition und Messung von hybrider Wertschöpfung. *IW Trends* 39: 1-20 und Lichtblau K, Arnold R. 2012. Smart Industry – Intelligente Industrie: Eine neue Betrachtungsweise der Industrie. Ergebnisse einer Studie der Institut der deutschen Wirtschaft Köln Consult GmbH für das Land Hessen, Initiative Industrieplatz Hessen, Neu-Isenburg.

Das Oslo-Handbuch (OECD & Eurostat 2018) unterscheidet zwei Arten von Produktinnovationen.⁵¹ Innovationen bei (1) Waren und (2) Dienstleistungen. Diese können schwer voneinander abzugrenzen sein, z.B. bei der Vermietung von langlebigen Gütern, der Bündelung von Waren und Dienstleistungen oder der Einbeziehung von Versicherungen in den Warenkauf. Wir haben bereits erörtert, wie wichtig Daten für praktisch jede Art von Produktinnovation sind, und ihr wohl wichtigster Einfluss dabei betrifft wohl die Art und Weise, wie solche Prozesse heute ablaufen. Insbesondere hat die Virtualisierung den Zeit- und Budgetaufwand für die Erstellung von Prototypen und die Erprobung und die Einführung neuer Produkte und Dienstleistungen verringert. Die Datenanforderungen für diese Prozesse sind relativ komplex und stützen sich in der Regel auf zahlreiche Datenquellen für einen sehr spezifischen Zweck. In einigen Fällen sind synthetische Daten erforderlich.⁵²

Der Nutzen von Daten ist besonders ausgeprägt, wenn diese in digitale Technologien integriert werden. Dadurch können sie zu einem wissensbasierten Produkt werden⁵³, das nicht nur Innovationen ermöglicht, sondern auch ein wichtiger Baustein der Innovationen selbst ist. Spiele, Musik und Video-Streaming-Dienste können als innovative, wissensbasierte Produkte betrachtet werden. Zwar können diese Dienste an sich als Innovationen angesehen werden, es ist jedoch klar, dass diese Innovationen ohne den Einsatz von Daten auf verschiedenen Ebenen nicht durchgeführt werden können. Beispielsweise stützen sich solche Dienste häufig auf innovative Datenformate sowie Komprimierungstechnologien und ähnliche Techniken, die ein positives Nutzererlebnis ermöglichen. Sie verfügen häufig über persönliche Empfehlungsmechanismen, die die

51 Es gelten die folgenden Definitionen: Waren: "Waren umfassen materielle Objekte und einige wissensbasierte Produkte (siehe unten), an denen Eigentumsrechte begründet werden können und deren Eigentum durch Markttransaktionen übertragen werden kann"; Dienstleistungen: "Dienstleistungen sind immaterielle Aktivitäten, die gleichzeitig produziert und konsumiert werden und die die Bedingungen (z.B. physische, psychische usw.) der Nutzer verändern. Das Engagement der Nutzer durch ihre Zeit, Verfügbarkeit, Aufmerksamkeit, Informationsweitergabe oder ihren Aufwand ist oft eine notwendige Bedingung, die zur Koproduktion von Dienstleistungen durch Nutzer und Unternehmen führt. Die Merkmale oder die praktische Erfahrung eines Dienstleistungsangebots können daher vom Input der Nutzer abhängen". OECD, Eurostat. 2018. *Oslo Manual 2018: Guidelines for Collecting, Reporting and Using Data on Innovation, 4th Edition, The Measurement of Scientific, Technological and Innovation Activities*. Paris and Luxembourg: OECD Publishing and Eurostat.

52 Siehe Abschnitt 2.5 für weitere Informationen.

53 Produkte zur Wissenserfassung sind im System of National Accounts definiert: "Produkte zur Wissenserfassung betreffen die Bereitstellung, Speicherung, Kommunikation und Verbreitung von Informationen, Beratung und Unterhaltung in einer Weise, dass die konsumierende Instanz wiederholt auf das Wissen zugreifen kann. Die Branchen, die die Produkte herstellen, sind diejenigen, die sich mit der Bereitstellung, Speicherung, Kommunikation und Verbreitung von Informationen, Beratung und Unterhaltung im weitesten Sinne dieser Begriffe befassen, einschließlich der Produktion von allgemeinen oder speziellen Informationen, Nachrichten, Beratungsberichte, Computerprogrammen, Filmen, Musik usw. Die Produkte dieser Branchen, für die Eigentumsrechte festgelegt werden können, werden oft auf physischen Objekten (sei es auf Papier oder auf elektronischen Medien) gespeichert, die wie gewöhnliche Waren gehandelt werden können. Sie weisen viele der Eigenschaften von Waren auf, da Eigentumsrechte an diesen Produkten begründet werden können und sie wiederholt verwendet werden können. Unabhängig davon, ob sie als Waren oder Dienstleistungen charakterisiert sind, haben diese Produkte die wesentliche gemeinsame Eigenschaft, dass sie von einer Einheit hergestellt und an eine andere geliefert werden können, wodurch Arbeitsteilung und die Entstehung von Märkten ermöglicht wird". European Commission, IMF, OECD, UN, World Bank. 2009. *System of National Accounts 2008*, European Commission, International Monetary Fund, Organisation for Economic Co-operation and Development, United Nations and the World Bank, New York, NY

Nutzerfreundlichkeit auf der Grundlage der getroffenen Entscheidungen oder der allgemeinen Präferenzen des Nutzers, die der Dienst erfasst hat, weiter verbessern. Schließlich erfordert die Strukturierung der meist großen Datenmengen über Titel, Genres, Interpreten, Verlage usw. eine innovative Datenverarbeitung. Während diese drei Beispiele nur die verschiedenen Rollen veranschaulichen können, die Daten als Teil der Innovationsaktivität im Rahmen und innerhalb der Bereitstellung digitaler Dienste spielen, zeigt sich, dass für eine erfolgreiche wissensbasierte Produktinnovation Daten in der Regel mit anderen technologischen Innovationen kombiniert werden müssen, um den gewünschten Erfolg zu erzielen.

Prozessinnovationen, wie sie im Oslo-Handbuch (OECD & Eurostat 2018) definiert sind, decken die gesamte Breite der Unternehmensprozesse ab.⁵⁴ Daten, die durch allgegenwärtige digitale Technologien und Praktiken in allen Bereichen moderner Unternehmen erfasst werden, liefern Informationen und ermöglichen letztlich solche Prozessinnovationen. Während die wichtigsten Daten aus den digitalen Technologien und Praktiken innerhalb des jeweiligen Unternehmens stammen dürften, kann es notwendig sein, diese um Daten anderer privater oder öffentlicher Einrichtungen zu ergänzen.⁵⁵

Typisches Beispiel: Wetterdaten helfen bei der Optimierung von Vorhersagen

Das Wetter hat einen grundlegenden Einfluss auf verschiedene Prozesse und Entscheidungen in Unternehmen. Wetterdaten sind vielfältig: Sie enthalten Zeit- und Ortsinformationen, die nicht nur für die Vergangenheit, sondern auch für die Zukunft vorliegen und die helfen, den zukünftigen Verlauf von Handlungen zu antizipieren. Daher kann der Zugang zu öffentlichen Wetterdaten dazu beitragen, Prozesse zu verbessern und in den verschiedenen Prozessphasen adäquat informierte Entscheidungen zu treffen. Dies gilt insbesondere für den Agrarsektor, wo die Ernten vom Wetter abhängen können. Agrarunternehmen können in der Voranpflanzungsphase für die Auswahl der Saatgutarten üblicherweise auf historische Daten und Prognosen zurückgreifen. In der Wachstumsphase können die Daten für Entscheidungen über Bewässerung und Düngung verwendet werden.⁵⁶

⁵⁴ Nach dem Oslo-Manual OECD, Eurostat. 2018. *Oslo Manual 2018: Guidelines for Collecting, Reporting and Using Data on Innovation, 4th Edition, The Measurement of Scientific, Technological and Innovation Activities*. Paris and Luxembourg: OECD Publishing and Eurostat. umfassen Prozessinnovationen: (1) Produktion von Waren oder Dienstleistungen; (2) Vertrieb und Logistik; (3) Marketing und Verkauf; (4) Informations- und Kommunikationssysteme; (5) Verwaltung und Management; (6) Produkt- und Geschäftsprozess-Entwicklung.

⁵⁵ Um den Zugang zu und den Austausch von Daten sowohl privater als auch öffentlicher Stellen zu erleichtern, hat die Europäische Kommission eine Verordnung über einen Rahmen für den freien Verkehr nicht-persönlicher Daten in der Europäischen Union (Verordnung (EU) 2018/1807) erlassen und die Richtlinie über offene Daten und die Weiterverwendung von Informationen des öffentlichen Sektors (Richtlinie (EU) 2019/1024) aktualisiert. Für weitere Informationen über diese legislativen Maßnahmen und ihre Auswirkungen siehe Barbero M, Cocoru D, Graux H, Hillebrand A, Linz F, et al. 2018b. Study on emerging issues of data ownership, interoperability, (re-)usability and access to data, and liability, Deloitte, Brussels und Barbero M, Bartz K, Linz F, Mauritz S, Wauters P, et al. 2018a. Study to support the review of Directive 2003/98/EC on the re-use of public sector information, Deloitte, Brussels.

⁵⁶ Die Erklärungen basieren hauptsächlich auf

Eine Reihe von Unternehmen bietet innovative Lösungen an, die diese Daten nutzen, um Geschäftsprozesse effizienter zu gestalten. Agrivi (<https://www.agrivi.com/en/>) bietet beispielsweise eine Betriebsführungssoftware zur Planung, Überwachung und Analyse der einzelnen Prozesse an. Die Software nutzt auch Wetterdaten, um Landwirte über optimale Zeiten für Spritz- und Schädlingsbekämpfungsmaßnahmen zu informieren.⁵⁷ Tracker.com stellt auch eine Betriebsführungssoftware zur Verfügung, die den Landwirten bei der Koordinierung von Prozessen hilft. Im Rahmen des Projekts "Big Data Agricultural Platform" beabsichtigt tracker.com, seine Software durch die Integration weiterer Datenquellen wie Wetter- und Satellitendaten zu erweitern. Insbesondere im Hinblick auf die Schädlingsbekämpfung will die Software die dokumentierten Daten mit Wetterdaten verknüpfen, um eine Prognose der Migration und der Entwicklung des Schädlings- und Unkrautbefalls zu ermöglichen.⁵⁸

Die Vision von Industrie 4.0 beruht auf der Erwartung substanzieller Prozessinnovationen, die auf der umfassenden Digitalisierung industrieller Wertschöpfungsketten basieren und diese praktisch in Wertschöpfungsnetzwerke verwandeln. Dies bedeutet nicht weniger als einen Paradigmenwechsel. Es wird erwartet, dass der so genannte digitale Zwilling dabei eine entscheidende Rolle spielen wird.⁵⁹ Der digitale Zwilling soll letztlich "selbst denkende Lieferketten" (Srai et al 2019) ermöglichen, die Produktionskapazitäten in Unternehmen und zwischen Unternehmen in Wertschöpfungsnetzwerken autonom bereitstellen. In der Regel wird ein direkter Ursache-Wirkungs-Zusammenhang angenommen, wenn der Zusammenhang zwischen der zunehmenden Digitalisierung – insbesondere der digitalen Zwillinge – und einem Paradigmenwechsel in der Lieferkette offensichtlich ist. In jüngster Zeit wird diese einseitige Ursache-Wirkungs-Annahme jedoch auch kritisch bewertet. Es scheint, dass die von Sensoren im Rahmen von Produktionsprozessen erfassten Daten in der Praxis nur selten wieder aufgegriffen werden. Daher ist *"die Verfügbarkeit technischer Daten nicht das Problem, sondern das Problem ist, die Zeit, die Werkzeuge und die Kompetenz für ihre Analyse bereitzustellen"* (Saudagar et al 2019). Srai et al (2019) weisen darauf hin, dass die größte Herausforderung für ein vollständig verbundenes (potenziell autonomes) Wertschöpfungsnetzwerk nicht die Daten sind, d.h. der digitale Zwilling selbst, sondern die Leistungsfähigkeit der Software hinter den Geschäftsprozessen ist. Tatsächlich sind *"die Herausforderungen, auf die man stößt [...], nicht rein technologischer Natur, sondern sie ergeben sich bei dem Versuch, den Daten einen Mehrwert zu verleihen und die gewonnenen Erkenntnisse in die Praxis umzusetzen [...]"* (Srai et al 2019). Ein wirklich innovativer

<https://www.ibm.com/blogs/insights-on-business/gbs-strategy/weather-means-business/> [Letzter Zugriff am 22.10.2019].

57 Siehe <https://www.agrivi.com/en/> [Letzter Zugriff am 22/10/2019] und https://zenodo.org/record/1406945/files/Report%20on%20successful%20innovation%20processes%20and%20best%20practices%20_20180831.pdf?download=1 [Last accessed 22/10/2019].

58 Siehe <https://www.trecker.com/index.php/eu-projekt-big-data-agrarplattform/> [Letzter Zugriff am 22/10/2019].

59 Für einen Überblick siehe Tao F, Zhang H, Liu A, Nee AY. 2018. Digital twin in industry: state-of-the-art. *IEEE Transactions on Industrial Informatics* 15: 2405-15.

Geschäftsprozess, der zu einem Unternehmen mit einem beträchtlichen Wettbewerbsvorteil passt, erfordert voraussichtlich wesentlich mehr als nur den Zugang zu Daten.

Aus der Literatur und den oben beschriebenen Fällen geht hervor, dass die Kontrolle über den Datenerfassungsprozess und damit ein vertieftes Wissen über den Inhalt und die Merkmale der Daten sehr wichtige Informationen sowohl für Produkt- als auch für Prozessinnovationen mit sich bringt. Beim Zugriff auf Daten Dritter sind die Möglichkeiten, die Merkmale der Daten vollständig zu bestimmen, vermutlich eingeschränkt. Dadurch können Innovationen erschwert werden, die einen Wettbewerbsvorteil mit sich bringen würden. Daher erscheint es plausibel, dass nur vier der sieben nationalen innovationspolitischen Richtlinien, die von Planes-Satorra and Paunov (2019) analysiert wurden, den Zugriff auf Daten über den auf der europäischen Ebene geltenden gesetzlichen Rahmen hinaus vorsehen.⁶⁰ Darüber hinaus bleiben die Vorschriften für den Datenzugriff in diesen Richtlinien entweder vage in ihren Konzepten, die auf die Verbesserung der rechtlichen Rahmenbedingungen (Deutschland), die Erforschung neuer Wege des Datenaustauschs (Großbritannien) und die Förderung offener Daten (China) abzielen, oder sie beschränken sich auf sektorspezifische Plattformen für die Sammlung und den Austausch von Daten (Frankreich).

Im innovationspolitischen Kontext können kontinuierlich in der realen Welt erfasste Daten schließlich dazu verwendet werden, die traditionellen amtlichen Statistiken zu erweitern, um bessere Erkenntnisse über die Innovationsaktivitäten und -ergebnisse in den verschiedenen Sektoren zu erhalten. Dies könnte künftige politische Maßnahmen maßgeblich beeinflussen und ihre Spezifität erhöhen.⁶¹

⁶⁰ Siehe Fußnote 55.

⁶¹ Für aktuelle Diskussionen über die Auswirkungen von (großen) Datenmengen auf die amtliche Statistik siehe z.B. Salgado D, Esteban E, Saldana S, Oancea B, Sakarovitch B, et al. 2018. *Estimation of population counts combining official data and aggregated mobile phone data*. Presented at European Conference on Quality in Official Statistics, Kraków und Wiengarten L, Zwick M. 2018. Neue digitale Daten in der amtlichen Statistik. *WISTA* 2017: 43-60.

Die wichtigsten Ergebnisse aus Kapitel 4

- *Big Data ist nicht zwangsläufig der beste Ansatz. Kleine Datensätze können je nach Anwendung effizienter sein als große Datensätze.*
- *Mit zunehmender Qualität und Quantität der verfügbaren Daten bestehen mehr Möglichkeiten, Beziehungen und Muster zu entdecken und neue Erkenntnisse zu gewinnen, was wiederum effizientere Prozesse, Produktverbesserungen und Dienstleistungsinnovationen ermöglicht.*
- *Skalenerträge und Verbundvorteile auf der Angebotsseite, die noch durch Netzwerkeffekte auf der Nachfrageseite und steigende Erträge durch die Wiederverwendung der Daten ergänzt werden, können zu datengetriebener Marktmacht und Marktkonzentration führen. Dies schließt jedoch nicht notwendigerweise den Wettbewerbsdruck für etablierte Unternehmen durch potenzielle neue Marktteilnehmer aus.*
- *Die Erhebung von und die ausschließliche Kontrolle über bestimmte Daten durch einige Unternehmen kann zu wettbewerbsrechtlichen Bedenken hinsichtlich des Datenzugangs (Marktzutrittsschranken) führen. Allerdings sind Daten als solche wertlos, wenn die Unternehmen nicht in der Lage sind, daraus Erkenntnisse zu gewinnen, die sie zur Verbesserung und/oder Monetarisierung ihrer Produkte und Dienstleistungen nutzen können.*
- *Die Datenqualität ist eine wichtige Ressource im Wettbewerb in der Datenökonomie.*
- *Aus der Perspektive von Industrie 4.0 ist Ko-opetition als eine Mischung aus Wettbewerb und Kooperation zwischen Unternehmen (z.B. Datenaustausch) das dominierende Form.*
- *(Persönliche) Daten erfüllen nicht die notwendigen Voraussetzungen, um als wesentliche Einrichtung zu gelten.*
- *Transparenz in Bezug auf die Datenverarbeitung und die Nutzungsbedingungen ermöglicht es Geschäftskunden und auch Endnutzern, ihre Ansprüche in Bezug auf Vertraulichkeit und Datenschutz gegen die Vorteile der Datenweitergabe abzuwägen.*
- *Transparenz in technischer Hinsicht kann dazu beitragen, Transaktionskosten zu reduzieren.*
- *Der Zugang zu offenen Daten ermöglicht es den Datennutzern, ein kooperatives Dienstleistungsangebot von mehreren Drittanbietern zu unterstützen.*
- *Transparenzverpflichtungen und der uneingeschränkte Zugang zu offenen Daten sind geeignete Instrumente, um den Wettbewerb in der Datenökonomie wirksam zu fördern.*

4 Wettbewerb in der Datenökonomie

In den vorhergehenden Abschnitten wurden die wirtschaftlichen Aspekte der Datennutzung untersucht. Offensichtlich haben Daten und die aus Daten gewonnenen Erkenntnisse einen erheblichen Einfluss auf Geschäftsprozesse und Rentabilität. Ein Unternehmen, das in der Lage ist, auf relevante Daten zuzugreifen, sie zu erfassen und zu seinem Vorteil zu nutzen, dürfte also wettbewerbsfähiger sein als ein Unternehmen, das dazu nicht in der Lage ist.

Während die Datenökonomie als ein sektorübergreifendes Phänomen an Dynamik gewinnt, sind Befürchtungen in Bezug auf die datengetriebene Marktmacht und den unterdrückten (zukünftigen) Wettbewerb aufgrund des unzureichenden Datenzugangs aufgekommen. Im Wesentlichen befürchten die Entscheidungsträger, dass sich der Datenzugang zu einer kritischen Eintrittsbarriere entwickeln könnte. Solche Bedenken sollten natürlich weiter untersucht werden, und deshalb befasst sich dieses Kapitel mit Fragen des Wettbewerbs in der Datenökonomie.

4.1 Rolle der Daten für den Wettbewerb

***Erkenntnisse:** Nicht für alle Arten von Anwendungen ist Big Data notwendig, da es effizienter sein kann, spezifische (kleine) Datensätze zu nutzen. Wenn Daten einen wertvollen Input darstellen, besteht die Tendenz, dass qualitativ hochwertige Daten besser verfügbar sind, was die Chancen auf bessere Ergebnisse – effizientere Prozesse, Produktverbesserungen und Dienstleistungsinnovationen – erhöht, als dies ohne solche Daten möglich wäre. Um die Vorteile des Datenzugangs ausschöpfen zu können, müssen Unternehmen erhebliche zusätzliche Anstrengungen und Kapazitäten aufbringen um im Wettbewerb zu bestehen.*

Um die Rolle von Daten für den Wettbewerb in der Datenökonomie verstehen zu können, ist es hilfreich, die frühere "Small Data" Ära und die neue "Big Data" Ära zu vergleichen, um einige qualitative Unterschiede hervorzuheben. In erster Linie spielen für die aktuellen Anwendungen kontinuierliche Datenflüsse eine zentralere Rolle als Datenbestände (Davenport et al 2012). Während die Datenmenge früher eine Spannweite von begrenzt bis groß umfasste, kann sie nun als sehr groß bezeichnet werden. In der jüngeren Vergangenheit gab es eine Knappheit an digitalen Daten und einen starken Bedarf an Zugang zu verschiedenen Datenquellen; eine Herausforderung bestand darin, die richtige Stichprobe zu erhalten. Heute erzeugen alle möglichen Geräte, Personen, Firmen und Institutionen alle Arten von digitalen Daten. Folglich ist die Ermittlung der am besten geeigneten Daten – das Durchsuchen der riesigen Menge kontinuierlich erfasster Daten – oft die zentrale Herausforderung für Unternehmen.

Bei den eingesetzten Informationssystemen gibt es eine enorme Entwicklung von geringer zu hoher Skalierbarkeit und Flexibilität im Zusammenhang mit Big Data. Da die Speicherkapazitäten steigen und somit noch größere Datenmengen erfasst werden können, verschiebt sich das Konzept von Big Data ständig. Was auch immer heute als Big Data gelten mag, wird in Zukunft möglicherweise nicht mehr dem Konzept entsprechen. Darüber hinaus definiert die Art der Daten auch, was mit "groß" gemeint ist, z.B. benötigt Video mehr Verarbeitungs- und Speicherkapazität als Text. Clickstream-Daten aus dem Web, Video-Streaming-Daten und Datenströme aus sozialen Medien erfordern individuelle Verarbeitungstechniken für Daten-Feeds.

Big Data bietet neue Möglichkeiten, auf der Erkennung von Mustern mit Hilfe von Methoden des maschinellen Lernens basieren, um verschiedenen Wirtschaftsakteuren neue und wertvolle Erkenntnisse zu liefern. Die Kombination verschiedener Datensätze, um daraus neue Informationen abzuleiten oder zu erkennen, die in einem bestimmten Kontext von wirtschaftlichem Wert sind, ist für die Datenökonomie von zentraler Bedeutung. Es ist jedoch nicht für alle Arten von Anwendungen notwendig, Big Data zu verwenden, da es effizienter sein kann, über spezifische (kleine) Datensätze verfügen zu können. Unabhängig von der Größe des Datensatzes wird seine Eignung für einen bestimmten Zweck beim Zugriff durch Dritte immer vom ursprünglichen Kontext und Zweck abhängig sein. Unternehmen mit Erstzugang zu Daten haben natürlich tendenziell mehr Einfluss auf den Kontext und den Zweck der ursprünglichen Datenerfassung. Das in Abschnitt 2.1 vorgestellte mehrschichtige Modell erklärt dies im Detail.

Die Erkenntnisse, die aus den Daten gewonnen werden können, verleihen bestimmten Datensätzen einen wirtschaftlichen Wert. Während dies offensichtlich zu sein scheint, erweist es sich in der Praxis als eine sehr komplexe Aufgabe (Feijóo et al 2016). Da Daten unterschiedlicher Herkunft und Art in verschiedenen Kontexten für einen unterschiedlichen Zweck verwendet werden können, hängt der spezifische Wert eines bestimmten Datensatzes auch von der spezifischen Aufgabe und Zielsetzung ab (Bründl et al 2015). Im Allgemeinen kann der wirtschaftliche Wert von Daten einerseits aus der Anbieterperspektive und andererseits aus der Nutzerperspektive bestimmt werden. Ein Indikator für den Wert von personenbezogenen Daten im Rahmen eines werbefinanzierten Geschäftsmodells ist beispielsweise der Werbeumsatz pro Nutzer (ARPU). Ansätze zur Bestimmung des Wertes sonstiger Datensätze sind stark vom jeweiligen Kontext abhängig, einschließlich des Geschäftsmodells, der Art der Daten, des Produkts, der Branche usw. Der Preis von Daten kann auch aus der Sicht eines (externen) Datenvermittlers bestimmt werden (Anthes 2015, Feijóo et al 2016, FTC 2014). Dabei ist der Preis pro Datensatz eine Funktion der Erhebungskosten, des aktuellen und zukünftigen Umsatzpotenzials, der wettbewerbsorientierten Nutzung der Informationen und der wirtschaftlichen Perspektiven sowie der Gesamtentwicklung der entsprechenden Branche.

Eine weit verbreitete (implizite) Annahme ist, dass Daten ziemlich homogen (konsistent) sind und dass es sich dabei um einen wichtigen Produktionsfaktor ähnlich wie Arbeit,

Kapital und Humankapital handelt (e.g., Farboodi & Veldkamp 2019, Jones & Tonetti 2019). Daten sind jedoch heterogen (vielfältig), so dass es schwierig ist, den rechtlichen Status und den ökonomischen Wert der verschiedenen Datentypen zu definieren. Dies zeigt sich auch in der Diskussion über Fragen des Datenbesitzes (Dosis & Sand-Zantman 2019, Duch-Brown et al 2017).

Der Datenzugang selbst ist nicht die Hauptvoraussetzung, um einen (datengetriebenen) Wettbewerbsvorteil zu erlangen, sondern es ist der Zugang zu Daten von angemessener Qualität. Qualität kann definiert werden als "*für die Nutzung durch Datenkonsumenten geeignet*" (Wang & Strong 1996). Das Ausmaß, in dem die Qualität der Daten bewertet werden muss, hängt in erster Linie vom Kontext der Datennutzung ab. Diese Bewertung kann sich auf mehr als 170 Dimensionen stützen, die u.a. von Wang and Strong (1996) beschrieben wurden. Nach Cichy and Rass (2019) gehören zu den wichtigsten objektiven Dimensionen der Datenqualität die Folgenden:

- **Vollständigkeit:** Der Grad, in dem Breite, Tiefe und Umfang der Daten ausreichend sind für die vorliegende Aufgabe.
- **Genauigkeit:** Das Ausmaß, in dem die Daten korrekt, zuverlässig, gültig und belegt sind.
- **Aktualität:** Das Ausmaß, in dem das jeweilige Alter der Daten für die vorliegende Aufgabe angemessen ist.
- **Konsistenz:** Das Ausmaß, in dem die Daten im gleichen Format vorliegen und mit früheren Daten kompatibel sind.
- **Zugänglichkeit:** Das Ausmaß, in dem Informationen verfügbar oder leicht und schnell abrufbar sind.

Die Datenqualität kann als eine wichtige Ressource im Wettbewerb in der Datenökonomie betrachtet werden, in der verschiedene Arten von Unternehmen systematisch verschiedene Arten von Daten(sätzen) sammeln, speichern, verarbeiten und nutzen. Wenn verschiedene Unternehmen auf der Grundlage datengestützter Produkte und Dienstleistungen miteinander konkurrieren, besteht die Tendenz, dass je höher die Qualität und Quantität der verfügbaren Daten ist, desto mehr Möglichkeiten bestehen, Beziehungen und Muster zu entdecken und neue Erkenntnisse zu gewinnen, die wiederum Produktverbesserungen und Dienstleistungsinnovationen ermöglichen (Junqué de Fortuny et al 2013, Martens 2016).

Bei der Beurteilung von Big Data aus der "Ressourcen-basierten Sicht des Unternehmens" (Barney 1991), die besagt, dass Big Data einzigartig, knapp, wertvoll und nicht substituierbar sein müssen, um einen Wettbewerbsvorteil zu ermöglichen, kommen Lambrecht and Tucker (2015) zu dem Schluss, dass 1) Big Data nicht einzigartig oder knapp sind, 2) Substitute existieren, 3) Big Data an sich wahrscheinlich nicht wertvoll sind und 4) es viele alternative Datenquellen gibt, die den Firmen zur Verfügung stehen. Dies entspricht auch Tucker and Wellford (2014), die argumentieren, dass Big Data einer von vielen Informationsinputs für die Dienstleistungen sind, die Online-

Unternehmen anbieten, wobei die meisten Firmen auch selbst relevante Informationen generieren. Sie kommen zu dem Schluss, dass Big Data weder ein Produkt im kartellrechtlichen Sinne noch die Art von Input sind, den Unternehmen von anderen erhalten müssen, um im Wettbewerb bestehen zu können. Dennoch können umfassende (Nutzer- und Nutzungs-)Daten ein wertvoller Input in der Datenökonomie sein, da sie 1) maßgeschneiderte Angebote, 2) personalisierte Empfehlungen und 3) gezielte Werbung ermöglichen (Fast et al 2019). Darüber hinaus gibt es auch Datenmarktplätze, die den Handel mit Datensätzen ermöglichen. Im Folgenden werden die wichtigsten Quellen digitaler Daten zusammengefasst: maßgeschneiderte Dienstleistungen, Empfehlungssysteme, gezielte Werbung und Datenmarktplätze.

Maßgeschneiderte Dienstleistungen können zu größerer Zufriedenheit und Kundentreue führen, die Umstellungskosten und Cross-Selling-Möglichkeiten erhöhen und die Zahlungsbereitschaft steigern (Ansari & Me-la 2003, Benlian 2015, Peppers et al 1999, Pine et al 1995, Tam & Ho 2006). Allerdings können datengetriebene Dienste auch Bedenken hinsichtlich des Daten- und Verbraucherschutzes aufwerfen, je nach Renommee und Betriebskontext des Unternehmens und der Art der Anpassung (Awad & Krishnan 2006, Chellappa & Sin 2005, Thirumalai & Sinha 2013).

Empfehlungssysteme können zu höheren Umsätzen führen als Systeme ohne entsprechende Empfehlungssysteme, vorausgesetzt, dass ausreichende, genaue und aktuelle Daten zur Verfügung stehen. Die Verwendung der Historie eines Kunden (z.B. aus Recherchen, Kaufabwicklungen, Servicenutzung) zur Bereitstellung passender Empfehlungen erfordert ein hohes Maß an Datenqualität und Datenquantität (O'Mahony et al 2006, Pipino et al 2002). Dies hängt außerdem von der Bereitschaft der Kunden ab, Daten offenzulegen, von ihrer Zahlungsbereitschaft und den Umsatz- und Cross-Selling-Möglichkeiten einer Firma (Adomavicius et al 2017, Hinz & Eckert 2010, Karwatzki et al 2017, Schafer et al 2001). Im Allgemeinen hängt die Qualität und Glaubwürdigkeit der Empfehlungen vom Zeitpunkt, der Bewertungsneutralität und der Transparenz gegenüber den Kunden sowie von deren Vertrauen in den datengetriebenen Dienst ab (Benbasat & Wang 2005, Ho et al 2011, Karwatzki et al 2017, Sinha & Swearingen 2002, Wang et al 2018).

Gezielte Werbung verbessert die Werbewirksamkeit in vielen Zusammenhängen. Sie ist abhängig von den Klickraten, den View-Through-Raten, der Kaufabsicht und der Kaufentscheidung der Nutzer (Bleier & Eisenbeiss 2015a, Bleier & Eisenbeiss 2015b, Goldfarb & Tucker 2011a, Goldfarb & Tucker 2011b, Kim et al 2019, Lambrecht & Tucker 2013, Tucker 2014). Wichtige Aspekte, die Werbetreibende berücksichtigen müssen, sind der Zeitpunkt und die Platzierung von Anzeigen, die Rechtfertigung der Werbung, das Kundenvertrauen, die empfundene Kontrollmöglichkeit, die

Transparenz und die Datenschutzbestimmungen (Aguirre et al 2015, Bleier & Eisenbeiss 2015a, Bleier & Eisenbeiss 2015b, Kim et al 2019, Samat et al 2017, Schumann et al 2014).

Es entstehen auch **Datenmarktplätze**, auf denen die Wirtschaftsdaten keinem der Beteiligten exklusiv zugeordnet werden und daher im Prinzip von jedem vermarktet werden können (Krämer & Wohlfarth 2018).

Im Wesentlichen zeigen all diese Beispiele, die in der aktuellen Literatur typischerweise als potentiell positive Ergebnisse des Datenzugriffs und der Datennutzung im geschäftlichen Kontext bezeichnet werden, dass Daten allein nur einen von vielen Bausteinen darstellen, die zu nachhaltigem Geschäftserfolg führen.

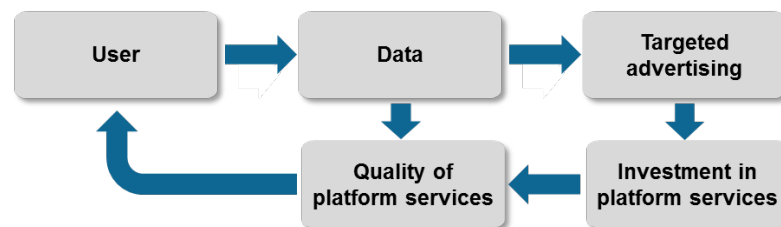
4.2 Datengetriebene Marktmacht und Markteintrittsbarrieren

Erkenntnisse: Skalen- und Verbundvorteile auf der Angebotsseite, die noch durch Netzwerkeffekte auf der Nachfrageseite und steigende Erträge durch die Wiederverwendung der Daten ergänzt werden, können zu datengetriebener Marktmacht und Marktkonzentration führen. Dies schließt jedoch nicht notwendigerweise den Wettbewerbsdruck für etablierte Unternehmen durch potenzielle neue Marktteilnehmer aus. Die Erhebung von und die ausschließliche Kontrolle über bestimmte Daten durch einige Unternehmen kann zu wettbewerbsrechtlichen Bedenken hinsichtlich des Datenzugangs (Marktzutrittsschranken) führen. Allerdings sind Daten als solche wertlos, wenn die Unternehmen nicht in der Lage sind, daraus Erkenntnisse zu gewinnen, die sie zur Verbesserung und/oder Monetarisierung ihrer Produkte und Dienstleistungen nutzen können. Daher ist die Datenqualität eine wichtige Ressource im Wettbewerb. Aus der Perspektive von Industrie 4.0 ist Kooperation die dominierende Form in der Datenökonomie.

Digitale Technologien verändern die wirtschaftliche Entscheidungsfindung und die Geschäftsmodelle, indem sie die Kosten für Recherche, Vervielfältigung, Transport, Nachverfolgung und Überprüfung verlagern (Goldfarb & Tucker 2019). Infolgedessen ermöglichen drastisch reduzierte Kosten für die Sammlung und Speicherung von Daten sowie Fortschritte bei den Analysetechniken eine höhere Effizienz von Unternehmen mit ganz unterschiedlichem Hintergrund (OECD 2018). Dies dürfte auch mit einer Veränderung der Wettbewerbslandschaft verbunden sein. Im Gegensatz zu den traditionellen Wirtschaftssektoren, die durch "Wettbewerb auf dem Markt" gekennzeichnet sind, umfasst die Datenökonomie ein breites Spektrum von Unternehmen mit datengesteuerten Produkten und Dienstleistungen, die sektorübergreifend agieren und daher in den meisten Fällen durch einen "Wettbewerb um den Markt" charakterisiert werden können (OECD 2016).

Die Wettbewerbseffekte der Datenerhebung werden oft als "positive Feedback-Schleifen" bezeichnet. Da die Datenerhebung zu erheblichen Verbesserungen der Dienstleistungen führen kann, können diese Dienstleistungen mehr Kunden anlocken, was wiederum die Unternehmen in die Lage versetzt, noch mehr Daten zu erheben, die wiederum zur Verbesserung ihrer Dienstleistungen genutzt werden können. Diese positiven Feedback-Schleifen, die durch die Datenerhebung ermöglicht werden, können es potenziellen Konkurrenzunternehmen erschweren, die gleiche Leistung wie der etablierte Anbieter zu erbringen. Dadurch wird dessen möglicherweise starke Marktposition weiter gestärkt wird, was zu einer Marktkonzentration oder sogar zu einer Marktbeherrschung führen kann.

Abbildung 4-1: Nutzer- und Monetarisierungs-Feedback-Schleifen



Quelle: OECD (2016).

Die positiven Feedback-Schleifen beziehen sich im Wesentlichen auf die durch Daten ermöglichten Skaleneffekte. Skaleneffekte in der Datenanalyse bedeuten, dass die mit der Erfassung und Verarbeitung eines *Datums* verbundenen Kosten (d.h. die Kosten pro Output-Einheit) mit größeren Datenmengen sinken. (Junqué de Fortuny et al 2013, Lewis & Rao 2015, Li et al 2016, Moore 1959). Dies beruht auf der Tatsache, dass die mit der Infrastruktur verbundenen Fixkosten hoch sind, während die Grenzkosten der Datenerhebung und -verarbeitung gegen (fast) Null tendieren. Die Größenordnung dieser Kostenvorteile kann auch davon abhängen, ob es sich um eine eigene Datenerhebung oder um Datenerhebung durch Dritte handelt.

Skaleneffekte sind in vielen Branchen durchaus üblich, und aus der Sicht des Wettbewerbs ist es entscheidend, ob stetig steigende Skalenerträge vorliegen oder ob diese abnehmen, wenn eine Firma eine ausreichend große Datenmenge erlangt hat (Lerner 2014). Der letztere Fall von abnehmenden Skalenerträgen impliziert, dass der Grenzwert von mehr Daten irgendwann abnimmt und somit die positiven Feedback-Schleifen begrenzt sind. Das Ausmaß der Skaleneffekte kann sich außerdem von einem datenbasierten Dienst zum anderen unterscheiden.

Auch können sich Verbundvorteile ergeben, wenn eine Firma eine Vielzahl von Dienstleistungen anbietet, die Daten sammeln, z.B. durch die Bündelung verschiedener datenbasierter Dienste (Panzar & Willig 1981). Die Verknüpfung dieser Daten kann den Unter-

nehmen mehr Erkenntnisse liefern und sie in die Lage versetzen, ihre Dienstleistungen zu verbessern und ihre Marktposition effektiver zu stärken als Unternehmen mit weniger Dienstleistungen. Je mehr Daten ein Unternehmen kombinieren kann, desto besser sind daher seine Chancen, Erkenntnisse zu gewinnen, die zur Stärkung seiner Marktposition genutzt werden könnten.

Skalen- und Verbundvorteile auf der Angebotsseite können durch direkte und indirekte Netzwerkeffekte auf der Nachfrageseite verstärkt werden. Direkte Netzwerkeffekte liegen vor, wenn der Wert eines Dienstes mit der Anzahl der Kunden steigt, d.h. wenn eine lineare Kostenerhöhung mit einem exponentiellen Anstieg des Wertes eines Dienstes verbunden ist (Katz & Shapiro 1985). Indirekte Netzwerkeffekte können vorliegen, wenn es (positive) Spillover-Effekte zwischen den verschiedenen Seiten eines (Plattform-) Dienstes gibt (Parker & van Alstyne 2005). Bei datengetriebenen indirekten Netzwerkeffekten kann ein Unternehmen die auf einer Seite der Plattform gesammelten Daten nicht nur zur Verbesserung seines Dienstes für diese Nutzergruppe nutzen, sondern auch zur Verbesserung seiner Dienstleistungen für andere (Nutzergruppen-) Seiten (Prüfer & Schottmüller 2017).

Die Wiederverwendung von Daten generiert zusätzliche Erträge in Bezug auf Umfang und Größe und stärkt die Marktposition des Dienstes. Skalen- und Verbundvorteile sowie direkte und indirekte Netzwerkeffekte können zu einer Marktkonzentration führen; dies schließt jedoch nicht notwendigerweise den Wettbewerbsdruck aus (potenzielle Konkurrenz durch neue Marktteilnehmer). Dies gilt insbesondere für datenbasierte Dienste, die als mehrseitige Plattformen agieren.

Ein wichtiger Faktor, der den Wettbewerbsvorteil von datengetriebenen Unternehmen einschränken könnte, ist die Aktualität der Daten. Da der Grad der Informationen, die Firmen aus Daten extrahieren können, typischerweise von ihrer Genauigkeit abhängt, könnte die Kompetenz, hochrelevante Daten zu sammeln und zu verarbeiten, von größerer Bedeutung sein als die bloße Größe eines Datensatzes, der auch veraltete Daten enthalten kann. Diese Einschränkung ist insbesondere für Dienstleistungen, die stark von der Aktualität der Daten abhängen, wie z.B. im Falle von gezielter Werbung, von erheblicher Bedeutung. Folglich müssen potenzielle Wettbewerber nicht unbedingt einen Datensatz erstellen, der der Größe des Datensatzes des etablierten Betreibers entspricht, sondern sie müssen vielmehr Wege finden, hochrelevante Daten zu sammeln, um einen "wettbewerbsfähigen Datensatz" zu erstellen, der nicht unbedingt der "gleiche Datensatz" ist (Schepp & Wambach 2016).

Erstklassige Daten, d.h. Daten der höchsten verfügbaren Qualität, können zu einem dauerhaften datengetriebenen Wettbewerbsvorteil führen (Krämer & Wohlfarth 2018). Kunden können von einem einzigen Diensteanbieter profitieren, z.B. wenn Netzwerkeffekte eine wichtige Rolle spielen. Darüber hinaus können die Möglichkeit der angebotsseitigen Substitution und die Notwendigkeit kontinuierlicher Innovationen zur Aufrecht-

erhaltung eines Wettbewerbsvorteils das Potenzial zur Ausnutzung einer marktbeherrschenden Stellung verringern.

Es kann nicht ausgeschlossen werden, dass die Erhebung von und die ausschließliche Kontrolle über bestimmte Daten durch einige Unternehmen zu wettbewerbsrechtlichen Bedenken Anlass gibt, doch sind Daten als solche wertlos, wenn die Unternehmen nicht in der Lage sind, Erkenntnisse zu gewinnen, die sie zur Verbesserung und/oder Monetarisierung ihrer Produkte und Dienstleistungen nutzen können. Wichtige Voraussetzungen für den Erfolg sind daher z.B. Experten wie Data Scientists und die Technologie, die zur Analyse der gesammelten Datenmengen eingesetzt wird. Darüber hinaus müssen datengetriebene Unternehmen auch mit abnehmenden Vorteilen bei steigenden Datenmengen rechnen (Li et al 2016), da sie mit einer mindestoptimalen Datenmenge umgehen müssen (Lewis & Rao 2015).⁶²

Es gibt Bedingungen, unter denen es für ein datengetriebenes Unternehmen möglich ist, seine Marktmacht auf der Grundlage der Sammlung und Nutzung von Daten aufrechtzuerhalten und zu stärken (Fast et al 2019). Während Netzwerkeffekte für sich genommen den Wettbewerb um den Markt fördern, kann ihre Kombination mit anderen Faktoren eine Monopolbildung begünstigen, die im Wesentlichen als Eintrittsbarriere zum Schutz des etablierten Unternehmens wirkt (Schweitzer 2019). Die Auswirkungen der Reputation einer Marke oder eines etablierten Unternehmens dürften große Unternehmen auf Kosten kleinerer begünstigen, wenn es um die Erhebung von Verbraucherdaten im Zusammenhang mit personalisierten Inhalten und Dienstleistungen geht (Chellappa & Sin 2005).

Marktbeherrschende Anbieter können den Wettbewerb ausschalten, indem sie die Markteintrittsbarrieren erhöhen (Haucap 2019, Haucap & Stühmeier 2016). Dies kann zu Zugangsproblemen für Konkurrenten und Neueinsteiger führen, die Zugang zu den von marktbeherrschenden Anbietern gesammelten Daten benötigen, um konkurrierende oder ergänzende Dienste anbieten zu können (Graef et al 2015). Eine Studie von Rubinfeld and Gal (2017) analysiert die verschiedenen Arten von Zugangsbarrieren, die den Zugang zu den verschiedenen Stufen der Daten-Wertschöpfungskette beschränken, und stellt fest, dass die spezifischen Merkmale von Big Data eine wesentliche Rolle spielen. Im Gegensatz dazu argumentiert Mahnke (2015), dass es nicht darauf ankommt, wie viele Daten ein Unternehmen hat, sondern dass es entscheidend ist, was es mit den Daten macht, wenn es seine Produkte und Dienstleistungen gestaltet und verbessert. Folglich muss bei der Wettbewerbsanalyse berücksichtigt werden, dass es zwar in einigen Datenmärkten einen starken Wettbewerb geben mag, dies aber nicht bedeutet, dass es in anderen keine Markteintrittsbarrieren gibt. Schweitzer (2019) argumentiert, dass der Zugang zu Daten nur dann über das Wettbewerbsrecht geregelt werden kann, wenn (1) Marktmacht auf den Primärmärkten oder (2) Marktmacht auf-

⁶² Wir gehen darauf in Abschnitt 2.5 näher ein.

grund der ausschließlichen Kontrolle über relevante Daten auf den Sekundärmärkten besteht.

Industrie 4.0 ist ein Paradigmenwechsel, bei dem traditionelle Wertschöpfungsketten nicht mehr die dominierende Organisationsstruktur sind (z.B. Firmen arbeiten auf (einer) bestimmten Ebene(n) der Wertschöpfung). Stattdessen gibt es Daten-Wertschöpfungsnetzwerke, in denen Unternehmen zunehmend gleichzeitig an unterschiedlichen Wertschöpfungsprozessen beteiligt sind.⁶³ Allerdings besteht bei Branchendaten derzeit ein Mangel an Rechtssicherheit bezüglich des Datenbesitzes und der (Wieder-)Verwendung von Daten. Aus der Wettbewerbsperspektive wird Ko-opetition dominieren, die das Verhältnis zwischen Unternehmen als eine Mischung aus Wettbewerb und Kooperation gleichzeitig charakterisiert.

Insgesamt gibt es in diesem frühen Entwicklungsstadium der europäischen Datenökonomie kein strukturelles Problem, das zu einem Marktversagen führt und eine sektorspezifische Regulierung notwendig machen könnte. Tatsächlich betreiben zahlreiche Unternehmen verschiedene Arten des Datenaustauschs auf der Grundlage bilateraler und multilateraler vertraglicher Vereinbarungen, wobei sie auch mit anderen Unternehmen im Wettbewerb stehen. Wie ein laufendes Forschungsprojekt der Europäischen Kommission zeigt,⁶⁴ " *tauscht ein beträchtlicher Anteil der Unternehmen in Europa und anderswo Daten aus, wobei der Austausch durch den Markt anstatt durch Regelungen zum Datenaustausch motiviert ist*". Politiker und Regulierungsbehörden, die die Datenökonomie stärken wollen, können sich jedoch auf die Förderung eines solchen Datenaustauschs mit einer geeigneten Standardisierung von Datenformaten, Datenportabilität, Interoperabilität und der Gewährleistung von Rechtssicherheit in Bezug auf Referenzarchitekturen konzentrieren. Dies wird in den Kapiteln 5 und 6 dieses Berichts näher erläutert. Tabelle 4-1 fasst die wichtigsten wettbewerbsrechtlichen Bedenken und die wichtigsten Beurteilungskriterien zusammen, die von Fall zu Fall geprüft werden müssen.

⁶³ Für eine Diskussion dieser Entwicklung siehe Kapitel 3.

⁶⁴ Siehe <http://datalandscape.eu/companies>.

Tabelle 4-1: Wettbewerbsrechtliche Bedenken und Beurteilungskriterien

Wichtigste wettbewerbsrechtliche Bedenken	Wichtigste Beurteilungskriterien
<ul style="list-style-type: none"> • Gibt es eine ausschließliche Kontrolle über bestimmte Daten, die eine erhebliche Zugangsbarriere darstellen? • Gibt es eine Übertragung von Marktmacht auf angrenzende Märkte? • Fehlt es an Wettbewerb bei nicht preisbezogenen Merkmalen wie dem Datenschutz? • Gibt es Informationsasymmetrien zwischen den Nutzern (d.h. Verbrauchern, Unternehmen) und dem Dienstanbieter, die sich aus einer bestimmten Marktposition mit Zugang zu einer (sehr) umfangreichen Menge und Vielfalt von (aktuellen) Daten ergeben? 	<ul style="list-style-type: none"> • Ökonomische Eigenschaften der relevanten Daten • Kontext und Zweck der Datennutzung • Zugang zu Datenquellen • Datenqualität • De-facto-Exklusivität von Daten • Möglichkeiten zur Reproduktion von Daten • Datenbasierte Größen- und Verbundvorteile • Kontextabhängigkeit der Daten in Bezug auf das entsprechende Produkt/die Dienstleistung • Wert der Daten, d.h. Informationswert

Quelle: WIK-Consult.

In der Datenökonomie dreht sich der Wettbewerb um den direkten Zugang zum (Geschäfts- und/oder End-) Nutzer über Schnittstellen (Application Programming Interfaces, APIs). Dies ermöglicht eine direkte Interaktionen zwischen Unternehmen und mit Kunden zur Anpassung von Diensten in (nahezu) Echtzeit. Dementsprechend werden bei Bedarf solche Schnittstellen entwickelt oder zur Verfügung gestellt, was zu einem marktgetriebenen Datenaustausch führt.

Eine häufige Frage in wettbewerbspolitischen Kreisen ist, ob eine weitere Verringerung der Barriere für den Datenaustausch den Wettbewerb in der Datenökonomie stärken würde. Bei der Beantwortung dieser Frage ist es hilfreich, zwei Fälle zu unterscheiden: (1) Erleichterung des Datenaustauschs durch den Endnutzer und (2) Erleichterung des Datenaustauschs durch Unternehmen.

Während eine künstliche Senkung der Zugangsbarriere für die gemeinsame Nutzung personenbezogener Daten weitere datenschutzrechtliche Bedenken aufwirft, kann sie auch das Vertrauen in die Praktiken des Datenaustauschs untergraben und als nachteilig für den Wettbewerb in der Datenökonomie angesehen werden. Der zweite Fall, die Erleichterung des Datenaustauschs durch Unternehmen, ist aus wettbewerbslicher Sicht relevant, da ungleiche Wettbewerbsbedingungen zwischen großen und kleinen Unternehmen den Wettbewerbsprozess behindern können, so dass eine Senkung der Zugangsbarriere für die gemeinsamen Nutzung von Daten hilfreich sein kann, wenn es um öffentliche Daten und Daten Dritter geht, nicht aber um vertrauliche Daten, z.B. Daten, die Geschäfts- und Betriebsgeheimnisse enthalten.

Eine weitere in wettbewerbspolitischen Kreisen häufig diskutierte Frage ist, ob (personenbezogene) Daten eine "wesentliche Einrichtung" sein können. Eine Doktrin über wesentliche Einrichtungen (Essential Facilities Doctrine, EFD) definiert, wann der/die

Eigentümer einer "wesentlichen" oder "Bottleneck"-Einrichtung verpflichtet ist/sind, Zugang zu dieser Einrichtung zu einem "angemessenen" Preis zu gewähren (OECD 1996). Das Konzept der "wesentlichen Einrichtungen" setzt zwei Märkte voraus, die oft als ein vorgelagerter und ein nachgelagerter Markt (z.B. zwei komplementäre Produkte/Dienstleistungen) bezeichnet werden. Typischerweise ist ein Unternehmen auf beiden Märkten tätig, während andere Unternehmen auf dem nachgelagerten Markt aktiv sind oder aktiv werden wollen. Ein nachgelagerter Wettbewerber möchte einen Input von dem betreffenden integrierten Unternehmen kaufen, wird aber abgelehnt. Ein EFD legt die Bedingungen fest, unter denen das integrierte Unternehmen den Zugang zu seinen Einrichtungen gewähren muss (Sidak & Lipsky 1999).

Um jedoch als "wesentliche" Einrichtung oder Infrastruktur gelten zu können, müssen (personenbezogene) Daten zwei notwendige Bedingungen erfüllen (Areeda & Hovorkamp 1988):

- 1) der Markteintritt in den komplementären Markt ist ohne Zugang zu dieser Einrichtung praktisch nicht möglich, und
- 2) ein Anbieter auf dem komplementären Markt kann diese Einrichtung nicht mit vertretbarem Aufwand reproduzieren; Substitute gibt es nicht.

Es ist erkennbar, dass (personenbezogene) Daten keine der beiden Bedingungen erfüllen. Was die erste Bedingung betrifft, so setzt sie voraus, dass ein Unternehmen kosteneffizienter ist als andere. Jede Firma in der Datenökonomie kann jedoch beliebige (personenbezogene) Daten zu oft vernachlässigbaren (nahezu Null) Grenzkosten sammeln, so dass es kein Monopol auf (personenbezogene) Daten gibt; die erste Bedingung ist also nicht erfüllt. Was die zweite Bedingung betrifft, so können (personenbezogene) Daten immer mit vertretbarem Aufwand dupliziert oder reproduziert werden, auch wenn dafür die Zustimmung des jeweiligen Endnutzers erforderlich ist. Die zweite Bedingung ist also auch nicht erfüllt. Insgesamt gibt es keine Fälle, in denen (personenbezogene) Daten als eine wesentliche Einrichtung angesehen werden könnten.

Um die aktuelle Debatte über die gemeinsame Nutzung von Daten mit entsprechender Expertise aufzuwerten, ist es wichtig, zu unterscheiden zwischen (1) Daten als (Plattform-)Dienstleistung, bei der es für etablierte Unternehmen Anreize für die gemeinsame Nutzung von Daten gibt, und (2) Daten als die ursprünglich erfassten Daten,⁶⁵ bei denen es für etablierte Unternehmen keine Anreize für die gemeinsame Nutzung gibt.⁶⁶

⁶⁵ Oftmals als "Rohdaten" bezeichnet.

⁶⁶ Wir gehen in Kapitel 5 dieses Berichts näher darauf ein.

4.3 Förderung des Wettbewerbs durch Transparenz und Zugang zu offenen Daten

***Erkenntnisse:** Transparenz in Bezug auf die Datenverwaltung und die Nutzungsbedingungen versetzt Geschäftskunden und Endverbraucher in die Lage, ihre Präferenzen in Bezug auf Vertraulichkeit und Datenschutz gegen die Vorteile der Datenweitergabe abzuwägen. Technische Transparenz kann dazu beitragen, die Transaktionskosten zu senken. Der Zugang zu offenen Daten ermöglicht es den Dateverwendern, ein gemeinsames Dienstleistungsangebot verschiedener Drittanbieter zu unterstützen. Transparenzverpflichtungen und der uneingeschränkte Zugang zu öffentlichen offenen Daten sind daher geeignete Instrumente, um den Wettbewerb in der Datenökonomie wirksam zu fördern.*

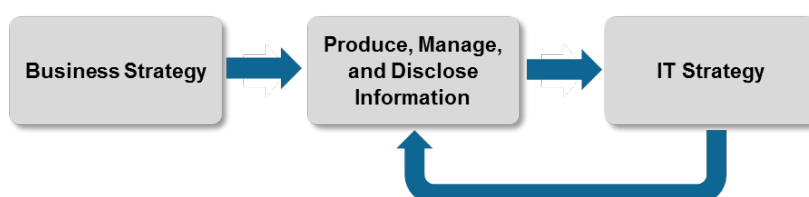
Die nächsten Schritte in Richtung einer einheitlichen europäischen Datenökonomie können eine sachgerechte Gestaltung der datenpolitischen Rahmenbedingungen beinhalten. Es steht außer Frage, dass Regelungen wichtig sind, da sie die möglichen Zugangspunkte zu Daten definieren und die Exklusivität und Pluralität in Bezug auf die Datenkontrolle regeln. Wenn es eine Informationsasymmetrie zwischen datengetriebenen Unternehmen und geschäftlichen Nutzern sowie Endnutzern hinsichtlich der Datenerhebung und -nutzung gibt, kann diese durch mehr Transparenz verringert werden.

Transparenzverpflichtungen versetzen Geschäftskunden und Endverbraucher in die Lage, ihre Präferenzen in Bezug auf Vertraulichkeit und Datenschutz gegen die Vorteile der Datenweitergabe abzuwägen. Darüber hinaus kann technische Transparenz dazu beitragen, die Transaktionskosten in der Datenökonomie zu senken (Tsai et al 2011). Daher können wirksame Transparenzverpflichtungen das Vertrauen in datenbasierte Dienste stärken und sie in die Lage versetzen, in Bezug auf die Dimensionen Vertraulichkeit und Datenschutz als Mittel zur Differenzierung von Produkten und Diensten in den Wettbewerb zu treten (Casadesus-Masanell & Hervás-Drane 2015).⁶⁷

⁶⁷ Es gibt jedoch drei Arten von Paradoxien, die derzeit Gegenstand weiterer Untersuchungen sind: 1) Paradoxon der Privatsphäre, Awad NF, Krishnan MS. 2006. The Personalization Privacy Paradox: An Empirical Evaluation of Information Transparency and the Willingness to Be Profiled Online for Personalization. *MIS Quarterly* 30: 13-28, Norberg PA, Horne DR, Horne DA. 2007. The Privacy Paradox: Personal Information Disclosure Intentions versus Behaviors. *Journal of Consumer Affairs* 41: 100-26, 2) Kontroll-Paradoxon, Brandimarte L, Acquisti A, Loewenstein G. 2012. Misplaced Confidences: Privacy and the Control Paradox. *Social Psychological and Personality Science* 4: 340-47 und 3) Transparenzparadoxon, Bernstein ES. 2012. The transparency paradox: A role for privacy in organizational learning and operational control. *Administrative Science Quarterly* 57: 181-216, Nissenbaum H. 2011. A Contextual Approach to Privacy Online. *Daedalus* 140: 32-48. Alle drei Paradoxe drehen sich um kognitive Voreingenommenheit bei der individuellen Entscheidungsfindung. Sie untergraben im Grunde die positiven Auswirkungen von Transparenzverpflichtungen und anderen Regulierungsansätzen. Solange diese Fragen nicht gelöst sind (z.B. die Auswirkungen dieser Phänomene), ist es nach wie vor wenig hilfreich, den Regulierungsbehörden und politischen Entscheidungsträgern vernünftige Empfehlungen zu geben.

Diesem Ansatz entspricht die geplante Richtlinie "New Deal for Consumers"⁶⁸ in Europa. Dementsprechend müssen digitale Plattformen wie z.B. Online-Marktplätze Transparenz hinsichtlich der wichtigsten Parameter ihres Rankings gewährleisten. Folglich müssen Online-Dienste auch darüber Auskunft geben, ob Dritte für ein besseres Ranking oder für die Aufnahme in die Ergebnislisten eine Gebühr zahlen. Darüber hinaus müssen digitale Plattformen wie z. B. Suchmaschinen nach der geplanten "Platform-to-Business"-Verordnung⁶⁹ bestimmte relevante Kriterien für ihr Ranking anderen Geschäftskunden und Verbrauchern zugänglich machen.

Abbildung 4-2: Transparenzstrategie



Quelle: Granados and Gupta (2013).

Im Vorgriff auf diese neuen Regelungen sollten Unternehmen der Datenökonomie eine Transparenzstrategie entwickeln (s. Abbildung 4–2), um Entscheidungen über die Weitergabe von Informationen innerhalb und außerhalb des Unternehmens treffen zu können, indem relevante Informationen selektiv anderen Geschäftskunden und Verbrauchern zugänglich gemacht werden (Granados & Gupta 2013). Insgesamt trägt ein angemessenes Maß an Transparenz dazu bei, einen fairen Wettbewerb zu begünstigen und gleichzeitig Geschäftskunden und Verbrauchern gerecht zu werden.

Tabelle 4-2: Grundsätze für offene Daten

Daten müssen komplett sein.	Alle Daten werden unter Beachtung der Datenschutz-, Sicherheits- oder Rechtebeschränkungen zur Verfügung gestellt.
Daten müssen primär sein.	Die Daten werden so veröffentlicht, wie sie an der jeweiligen Informationsquelle erhoben wurden, mit der feinstmöglichen Granularität, nicht in aggregierter oder modifizierter Form.
Daten müssen aktuell sein.	Die Daten werden so schnell wie nötig zur Verfügung gestellt, um ihren Wert zu erhalten.
Daten müssen zugänglich sein.	Die Daten stehen den unterschiedlichsten Nutzergruppen für die unterschiedlichsten Zwecke zur Verfügung.
Daten müssen maschinen-	Die Daten sind sinnvoll strukturiert, um ihre automatisierte

⁶⁸ Proposal for a Directive as regards better enforcement and modernization of EU consumer protection rules, COM(2018)0185 final-2018/090 (COD).

⁶⁹ Proposal for a Regulation on promoting fairness and transparency for business users of online intermediation services, COM(2018)238/974102.

verarbeitbar sein.	Verarbeitung zu ermöglichen.
Der Zugang muss diskriminierungsfrei sein.	Die Daten sind für jedermann verfügbar, ohne dass eine Registrierung erforderlich ist.
Die Datenformate dürfen nicht urheberrechtlich geschützt sein.	Die Daten stehen in einem Format zur Verfügung, über das kein Unternehmen die ausschließliche Kontrolle hat.
Daten müssen lizenzfrei sein.	Die Daten unterliegen keinen Bestimmungen über Urheberrechte, Patente, Marken oder Geschäftsgeheimnisse. Angemessene Einschränkungen der Vertraulichkeit, der Sicherheit und der Berechtigungen können gemäß anderer Vorschriften zulässig sein.
Die Einhaltung muss überprüfbar sein.	Es muss eine Ansprechperson für die Nutzer der Daten und für Beschwerden über die Nichteinhaltung der Regeln benannt werden, und eine weitere Instanz muss befugt sein zu entscheiden, ob die Regeln eingehalten wurden.

Quelle: <http://www.opengovdata.org>.

Ein weiterer Ansatz zur Förderung des Wettbewerbs besteht darin, offene Daten zu fördern, was bedeutet, dass ein öffentlicher Datenanbieter seine Daten einem offenen Kreis von Datennutzern zur Verfügung stellt, um die Entwicklung eines Ökosystems mit mehreren Drittanbietern zu fördern (Argenton & Prüfer 2012). Öffentliche Organisationen können den Zugang zu offenen Daten ermöglichen und sicherstellen, um den Wettbewerb in der Datenökonomie wirksam zu fördern.

Die wichtigsten Ergebnisse von Kapitel 5

- *Finanzdatenanbieter, Datenbroker und Online-Aggregatoren sind die wichtigsten Wirtschaftsakteure, die den Handel mit Daten auf den Datenmärkten ermöglichen. Ihre Produkte und Dienstleistungen können die Wahlmöglichkeiten der Kunden erhöhen und die Such- und Transaktionskosten senken. Sie werfen aber auch Fragen hinsichtlich des Datenschutzes auf.*
- *Der horizontale Datenaustausch erfolgt meist auf der Grundlage bilateraler vertraglicher Vereinbarungen und gemeinsamer Initiativen zwischen Unternehmen. Die Wertschöpfung beim horizontalen Datenaustausch entsteht beim Überführen von Daten in Informationen (kontextabhängig).*
- *Beim vertikalen Datenpooling werden modernste Technologien und Standards eingesetzt, um beim Übergang von der Datenerfassung und Datenverarbeitung durch die Gewinnung von Erkenntnissen über die Informationsbereitstellung und Informationsnutzung einen Mehrwert zu schaffen. Datenpooling ermöglicht daher ein hohes Maß an Innovation, insbesondere im Kontext von Industrie 4.0.*
- *Die allgemeinen Wohlfahrtseffekte sind je nach der spezifischen Ausgestaltung der Datenaustauscharchitektur unterschiedlich.*
- *Datenportabilität und Interoperabilität (Kompatibilität) sind für Innovationen von zentraler Bedeutung.*
- *Während sich die Datenportabilität meist auf offene Spezifikationen bezieht, ist die Interoperabilität mit formalisierten Standards verbunden.*
- *Für die Datenportabilität fehlen derzeit einheitliche technische Standards hinsichtlich der Datenformate und Verarbeitungsprotokolle für die Datenextraktion und -implementierung, so dass das Prinzip der Datenportabilität in der Praxis derzeit wenig wirksam ist.*
- *Interoperabilität kann dazu beitragen, die Macht der Netzwerkeffekte (d.h. die Marktmacht der etablierten Unternehmen) zu schwächen, aber auf Kosten der wirtschaftlichen Effizienz.*
- *Je nach ihrer spezifischen Ausgestaltung können beide Instrumente zu einem geringeren Maß an Vertraulichkeit und Datenschutz führen.*

5 Ansätze zur gemeinsamen Nutzung von Daten

Ein Schlüsselement des gängigen Narrativs über die Datenökonomie ist ein scheinbarer Mangel an Datenaustausch. Dieses Kapitel befasst sich mit den verschiedenen Arten des Datenaustauschs, die bereits stattfinden.

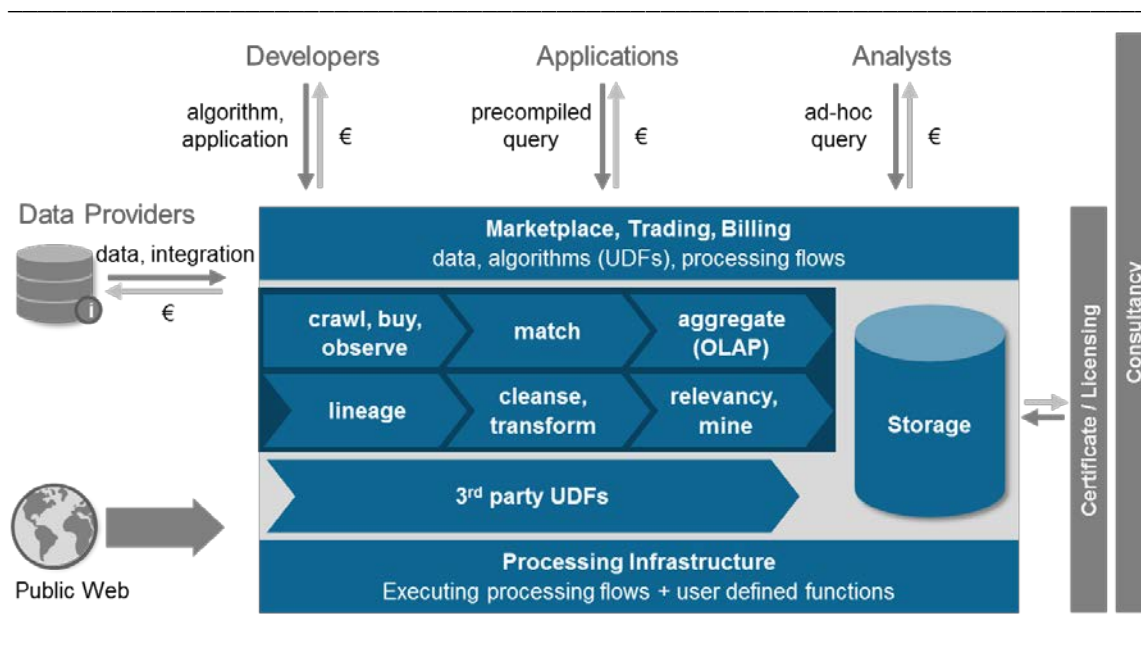
5.1 Datenmärkte und Handel mit Daten

***Erkenntnisse:** Anbieter von Finanzdaten und Kreditrating-Agenturen, die Nutzerbewertungen vornehmen, Datenhändler, die riesige Datensätze über einzelne Verbraucher zusammenstellen, und Online-Aggregatoren, die öffentlich zugängliche Daten zur Erstellung von Verbraucherprofilen auswerten, stellen die wichtigsten Akteure dar, die Datenmärkte und den Handel mit Daten ermöglichen. Sie bieten verschiedene Produkte und Dienstleistungen an, erhöhen die Wahlmöglichkeit der Kunden und senken die Such- und Transaktionskosten. Sie werfen aber auch Fragen hinsichtlich des Datenschutzes auf.*

Es gibt im Wesentlichen drei Arten von Datenanbietern, die auf kommerziellen Datenmärkten tätig sind (Bergemann & Bonatti 2019). Je nach Herkunft der Daten kann zwischen ihnen unterschieden werden: 1) Anbieter von Finanzdaten (z.B. Bloomberg, Thomson Reuters) und Kreditrating-Agenturen (Moody's, Standard & Poors, Fitch), die die Nutzerbewertungen vornehmen; 2) Datenhändlern (z.B. Acxiom, Palantir, Lexis-Nexis), die riesige Datensätze über einzelne Verbraucher zusammenstellen; und 3) Online-Aggregatoren (z.B. Intelius, Spokeo), die öffentlich zugängliche Daten auswerten, um Verbraucherprofile zu erstellen. Laut einer Studie des Pew Research Center können sich Verbraucher typischerweise auf drei Arten am Datenhandel beteiligen: 1) Daten gegen Dienstleistungen; 2) Daten gegen Bezahlung; 3) Daten als Spende.⁷⁰

⁷⁰ Umfrage des Pew Research Center unter amerikanischen Verbrauchern zur gemeinsamen Nutzung von Daten: <https://www.pewinternet.org/2016/01/14/privacy-and-information-sharing/>

Abbildung 5-1: Daten-Marktplatz



Quelle: Muschalle et al (2012).

Abbildung 5-1 veranschaulicht, wie ein Datenmarktplatz funktioniert. Wie dieses Beispiel der Integration öffentlicher Webdaten mit anderen Datenquellen zeigt, enthält es Komponenten für die Datenextraktion, die Datentransformation und das Laden von Daten sowie Metadaten Speicher, die Daten und Algorithmen beschreiben (Muschalle et al 2012). Der Datenmarktplatz verfügt über Schnittstellen für die Datenintegration, Methoden zur Optimierung und Komponenten für den Handel und die Abrechnung mit Dritten. Der Betreiber des Datenmarktplatzes erhält von seinen Kunden (z.B. Entwickler, Analysten, Anwender) eine monetäre Vergütung, wobei er sich auf eine skalierbare Infrastruktur für die Verarbeitung und Indizierung der Daten stützt.

Im Rahmen einer Untersuchung der US Federal Trade Commission (FTC 2014) zu den kommerziellen Datenmärkten wurden neun große Datenhändler, ihre Geschäftsmodelle und Praktiken analysiert. Die Untersuchung ergab, dass Datenhändler Informationen über Nutzer aus Quellen wie öffentlichen Posts, Online-Käufen, Browserverlauf und Garantiekarten zusammenstellen. Während sie Daten hauptsächlich aus allgemein zugänglichen Regierungsquellen sammeln, nutzen sie auch soziale Medien, Blogs und andere kommerzielle Datenquellen. Diese Datenhändler bieten im Wesentlichen drei Arten von Produkten und Dienstleistungen an:

- Online- und Offline-Marketing-Analysen;
- Risikominimierung, z.B. Scoring, Überprüfung von Identitäten, Erkennung von Betrug;
- Personensuche, z.B. Competitive Intelligence Services, Auffinden alter Freunde.

Datenhändler bieten den Nutzern eine größere Auswahl an und sie senken die Such- und Transaktionskosten; ihr Geschäftsmodell wirft jedoch Bedenken hinsichtlich des Datenschutzes auf. Während zahlreiche Organisationen über Daten verfügen, die für Einzelpersonen nicht als schlüssig anzusehen sind, können sie für große (globale) Organisationen nützlich sein, um komplexe Probleme aus gesellschaftlicher Sicht zu lösen. Aufgrund strategischer, rechtlicher und politischer Bedenken wird der Datenaustausch zögerlich betrieben. In einem geschützten Datenmarkt sind die ausgetauschten Daten sicher und stehen in granularer oder aggregierter Form auf der Grundlage der spezifischen Anforderungen der Anbieter und Kunden zur Verfügung. Der Datenhandel wird zwar als wesentlich für die Datenökonomie angesehen, doch gibt es zwischen den ursprünglich erfassten und den an Kunden verkauften Daten mehrere Zwischenstufen von Datenhändlern (die die Daten zunehmend verarbeiten und verfeinern), was es für den Nutzer schwierig macht, die von den Kunden eines Datenhändlers verwendeten Daten zurückzuverfolgen (Duch-Brown et al 2017).

Die meisten Datenhändler, wie z.B. Palantir, bieten verschiedene Arten von Produkten und Dienstleistungen an, z.B. erfasste Daten, Analyseergebnisse, Datenanalysesoftware und -kurse sowie Beratungsdienste. Aus wirtschaftlicher Sicht sind Datenhändler in der Regel mit hohen Fixkosten für die Erstellung ihrer datenbasierten Produkte und Dienstleistungen und niedrigen Grenzkosten pro Dateneinheit konfrontiert. Deshalb gibt es auf den Datenmärkten Skalen- und Verbundvorteile. Diese Kostenstruktur führt zwar nicht zu einer kostenbasierten Preisgestaltung, aber sie führt zu einer Preisgestaltung nach dem Wert für den Kunden, z.B. bei der Versionierung von Informationsgütern (Shapiro & Varian 1999). Die Preispolitik der Datenhändler auf den verschiedenen Datenmärkten ist weitgehend unbekannt, kann aber Folgendes umfassen: kostenlose Daten, die von öffentlichen Stellen bezogen werden, wie z.B. statistische Daten; nutzungsabhängige Preisgestaltung; Paketpreise; Abonnements; zweistufige Tarife, die aus einer festen Gebühr und einer variablen Gebühr pro verkaufter Einheit bestehen; und Freemium, eine Kombination aus einer kostenlosen Version mit Basisdiensten und einer kostenpflichtigen Version mit Premiumdiensten.

5.2 Horizontaler Datenaustausch und vertikales Datenpooling

***Erkenntnisse:** Der horizontale Datenaustausch erfolgt häufig auf der Grundlage bilateraler vertraglicher Vereinbarungen und Initiativen zwischen Unternehmen, bei denen die Wertschöpfung bei der Verarbeitung von Daten zu Informationen erfolgt (kontextabhängig). Im Gegensatz dazu nutzt das vertikale Datenpooling modernste Technologien und Standards, um einen Mehrwert auf dem Weg von der Datenerfassung, der Datenverarbeitung und dem Erlangen von Erkenntnissen bis hin zur Bereitstellung und Nutzung von Informationen zu generieren. Datenpooling ermöglicht daher ein hohes Maß an Innovation – insbesondere im Kontext von Industrie 4.0.*

Der häufigste Weg zu einem **horizontalen** Datentransfer im Hinblick auf die **gemeinsame Nutzung von Daten** sind bilaterale vertragliche Vereinbarungen und gemeinsame Initiativen von Unternehmen, wie die Studie von Barbero et al (2018b) zeigt. Eine Marktforschungsstudie von Pauer et al (2018) über den Datenaustausch zwischen verschiedenen Unternehmen aus allen Branchen zeigt, dass 75% der Führungskräfte die Möglichkeiten zur Verbesserung der Kundenbeziehungen, der Kundenkontakte und der Dienstleistungen als sehr gut einschätzen. Mehr als zwei Drittel betrachten darüber hinaus die Optimierung sowohl der Unternehmensprozesse als auch der Lieferketten als sehr große Chance. Eine von der Europäischen Kommission in Auftrag gegebene Studie über die gemeinsame Nutzung von Daten zwischen Unternehmen (B2B) kommt zu dem Schluss, dass (1) Unternehmen Daten bereits gemeinsam nutzen und wiederverwenden; (2) ihr Anteil in naher Zukunft steigen wird; (3) dies die Geschäftschancen erhöht und die interne Effizienz verbessert; (4) Investitionen in den Echtzeit-Zugriff auf Daten oder Lokalisierungsdaten sich positiv auf das Geschäftsergebnis eines Unternehmens auswirken können; (5) die meisten Datenanbieter und -nutzer Daten anscheinend innerhalb ihres eigenen Unternehmenssektors gemeinsam nutzen und wiederverwenden; (6) Dateninhaber nur einen kleinen Teil ihrer verfügbaren Daten teilen; (7) technische und rechtliche Hindernisse die gemeinsame Nutzung von B2B-Daten behindern, während die Verweigerung des Zugriffs ein übliches Hindernis für Unternehmen ist, die Daten wiederverwenden; und (8) Vertrauen und einfache Verfahren Unternehmen helfen würden, die Daten gemeinsam zu nutzen (Arnaut et al 2018).

In Bezug auf den horizontalen Datenaustausch und seine Wohlfahrtseffekte zeigen Jentzsch et al (2013), dass die Anreize für den Datenaustausch zwischen Unternehmen von der Art der Kundendaten und der Kundenvariabilität abhängen. Die Anreize zur gemeinsamen Nutzung von Daten sind stärker, wenn die Kunden sich relativ ähnlich sind. Für die Konsumentenrente ist die gemeinsame Nutzung von Kundendaten mit großer Wahrscheinlichkeit schädlich, während ihr Einfluss auf die soziale Wohlfahrt positiv sein kann. Wenn konkurrierende Unternehmen dazu gebracht werden sollen, ihre kundenspezifischen Daten gemeinsam zu nutzen, kann eine ausreichende Unter-

nehmensasymmetrie erforderlich sein (Liu & Serfes 2006). Beispielsweise kann ein Unternehmen mit geringerer Qualität seine Kundendatenbank an ein Unternehmen mit hoher Qualität verkaufen. Das Unternehmen mit hoher Qualität wird seine Daten jedoch niemals an den Konkurrenten mit niedriger Qualität verkaufen.

Die kontinuierliche oder gelegentliche Übertragung von Daten zwischen Unternehmen in einer horizontalen Beziehung zu bestimmten Zwecken ermöglicht hauptsächlich die Steigerung der Effizienz von Geschäftsprozessen, Produktverbesserungen, Dienstleistungsinnovationen, die Anpassung von Dienstleistungen und Empfehlungen. Koopetition auf der Grundlage von vertraglichen Vereinbarungen über die horizontale gemeinsame Nutzung von Daten wird voraussichtlich zu einer Steigerung der sozialen Wohlfahrt führen. Marktbedingte Beschränkungen von horizontalen Vereinbarungen über die gemeinsame Nutzung von Daten hängen davon ab, ob der Wert der geteilten Daten aus einem *Datum* oder eher aus seiner Agglomeration und der anschließenden Analyse stammt. Darüber hinaus kann das Wettbewerbsrecht die (Daten-) Zusammenarbeit zwischen (konkurrierenden) Unternehmen einschränken, z.B. in Bezug auf Verkäufe, Kartelle, datenbasierte Marktzutrittschranken.⁷¹ Die Wertschöpfung beim horizontalen Datenaustausch erfolgt also während der Entwicklung von Daten zu (kontextabhängigen) Informationen.

Im Gegensatz dazu schafft der **vertikale** Datenaustausch im Sinne einer **Bündelung von Daten** einen Mehrwert aus (1) der Datenerfassung, (2) der Datenverarbeitung, (3) der Gewinnung von Erkenntnissen, (4) der Informationsbereitstellung und (5) der Nutzung von Informationen. Der Aufbau von Datenpools zum Austausch und zur Nutzung von Daten durch Unternehmen in einer vertikalen Beziehung spielt eine zunehmend wichtige Rolle in der Datenökonomie. Drei Arten von Wirtschaftsakteuren können einen berechtigten Anspruch auf die daraus resultierenden Erträge haben: 1) die betroffene Person, um die es bei den Daten geht; 2) der Datenerfasser, der die Kosten für den Schutz der Daten trägt; und 3) der Datenverarbeiter, der die Daten analysiert, um daraus Erkenntnisse zu gewinnen (Carrière-Swallow & Haksar 2019). Jeder dieser Akteure hat seine eigenen Interessen, und seine Entscheidungen hinsichtlich der Anforderungen an die Datenbündelung werden sich auf die anderen auswirken.

Ein vielversprechendes Beispiel für eine Allianz privater Firmen zur Implementierung eines Datenpools ist der Industrial Data Space (Otto et al 2016, Otto et al 2018). Dieses Projekt zielt auf die Standardisierung von Internet of Things-Anwendungen ab und besteht im Wesentlichen aus fünf Schichten, die jeweils die für ein effektives Datenpooling notwendigen Elemente spezifizieren:

⁷¹ Es gibt auch rechtliche Ausnahmen wie z.B. die EU-Gruppenfreistellungsverordnungen für horizontale und vertikale Vereinbarungen.

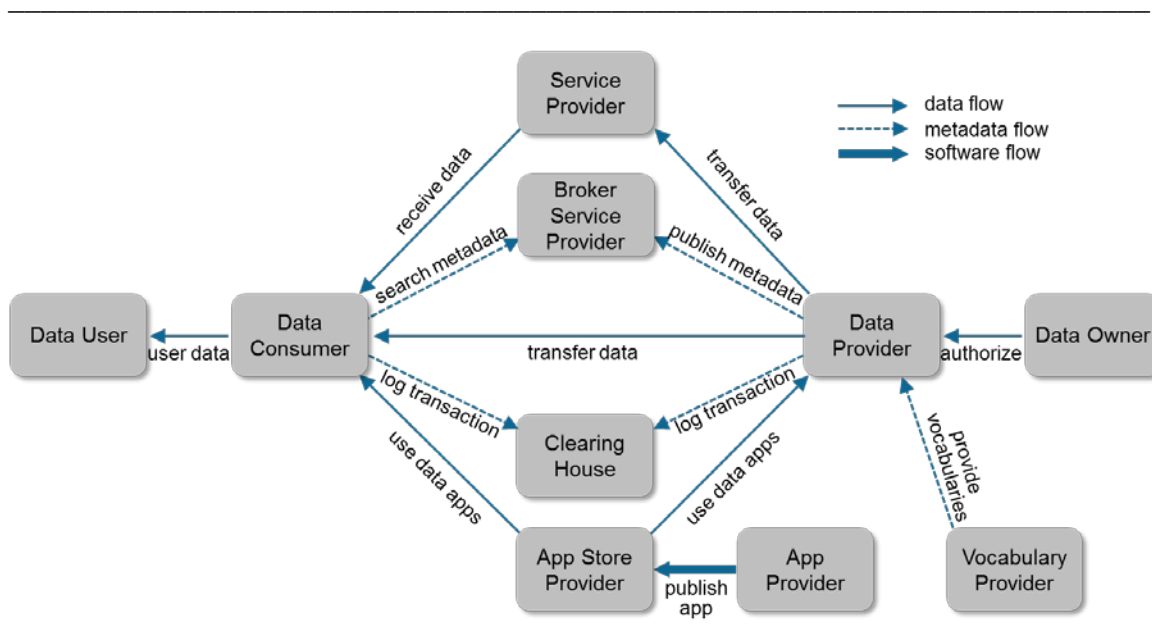
Tabelle 5-1: Schichten des Industrial Data Space (IDS)

Schicht	Beschreibung
System	Ordnet die Rollen der Geschäftsschicht einer Daten- und Service-Architektur zu; ein technischer Kern, der auf drei Hauptkomponenten basiert: Konnektor, Vermittler und App Store.
Informationen	Legt die Kompatibilität und Interoperabilität des Datenpools fest.
Prozess	Spezifiziert die Interaktionen zwischen verschiedenen Komponenten; Bereitstellung von Daten, Austausch von Daten, Veröffentlichung und Verwendung von Datenanwendungen.
Funktion	Definiert die zu implementierenden funktionalen Anforderungen und Merkmale.
Business	Legt die verschiedenen Rollen fest, die die Teilnehmer übernehmen können.

Quelle: Otto et al (2018).

Im Allgemeinen sieht das IDS-Projekt aus der Geschäftsperspektive mehrere Rollen vor: Datenbesitzer, Datenanbieter, Datenkonsument, Datennutzer, Vermittlungsdienstleister, Clearingstelle, Identitätsgeber, App-Store, App-Provider, Semantikanbieter, Softwareanbieter, Service-Provider sowie eine Zertifizierungsstelle und Evaluierungsstelle. Je nach Status empfängt und/oder sendet jede dieser Rollen (Inhalts-) Datenflüsse, Metadatenflüsse und Softwareflüsse. Abbildung 5-2 veranschaulicht diese Rollen und ihre Interaktionen auf der Geschäftsebene.

Abbildung 5-2: Rollen und Interaktionen (Geschäftsschicht)

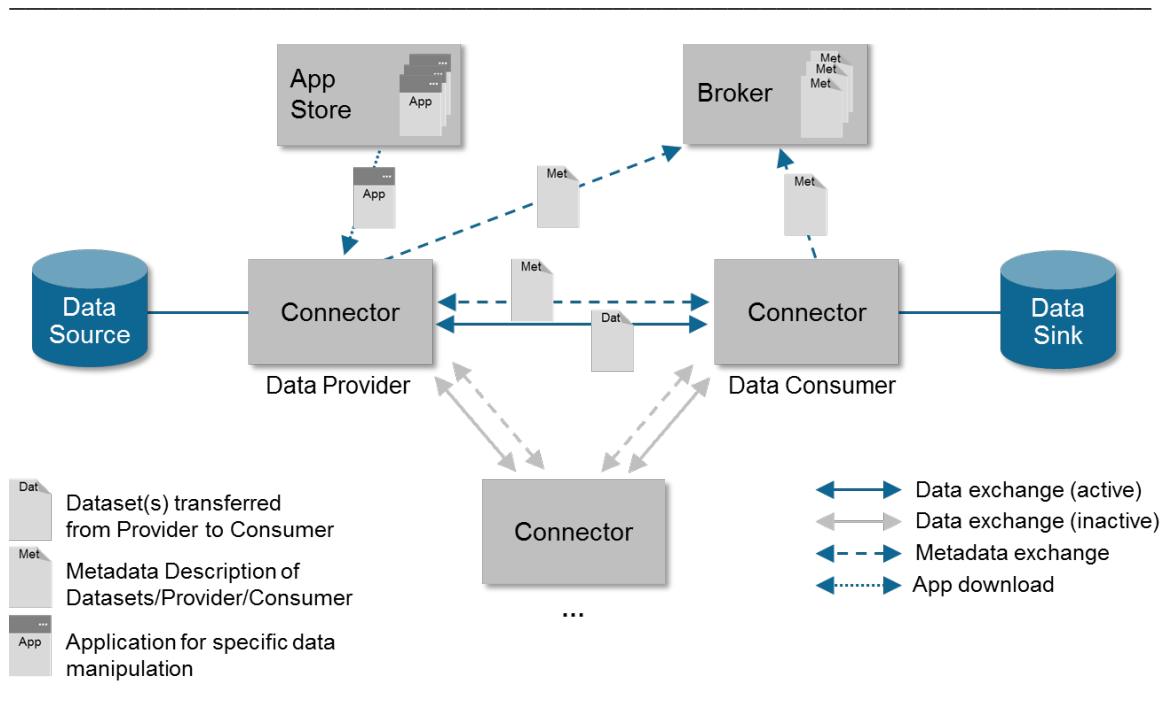


Quelle: Otto et al (2018).

Auf der Funktionsschicht werden die einzelnen Anforderungen in sechs funktionale Einheiten gruppiert: Vertrauen, Sicherheit und Datenhoheit, Datenökosystem, standardi-

sierte Interoperabilität, wertschöpfende Anwendungen (Datenverarbeitungssoftware) und Datenmärkte. Schließlich bildet das Zusammenspiel der technischen Komponenten, einschließlich der Kontrolle der Datennutzung, die Systemschicht, wie in Abbildung 5-3 dargestellt.

Abbildung 5-3: Zusammenspiel technischer Komponenten einschließlich der Kontrolle der Datennutzung (Systemschicht)

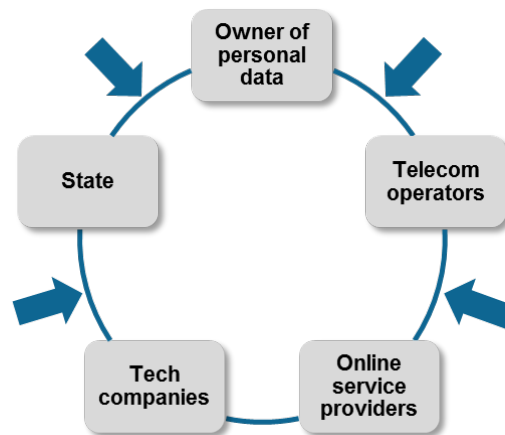


Quelle: Otto et al (2018).

Insgesamt "ist der Industrial Data Space ein virtueller Datenraum, der bestehende Standards und Technologien sowie akzeptierte Governance-Modelle für die Datenökonomie nutzt, um den sicheren und standardisierten Austausch und die einfache Verknüpfung von Daten in einem vertrauenswürdigen Geschäftsökosystem zu erleichtern. Er bietet damit eine Grundlage für intelligente Serviceszenarien und innovative unternehmensübergreifende Geschäftsprozesse, während gleichzeitig die Datenhoheit für die beteiligten Dateneigentümer gewährleistet ist" (Otto et al 2018).

Ein Beispiel für eine öffentliche Einrichtung, die den Datenpool-Anbieter repräsentiert, ist das Projekt German Data Trust (Lind & Suckfüll 2013). Die Projektarchitektur strebt eine ordnungsgemäße Behandlung und Nutzung von personenbezogenen Daten durch einen deutschen Datenpool-Anbieter an, die auf der Idee der GEMA, einem bekannten Datenpool-Anbieter in der Musikindustrie, basiert. Die führenden Rollen in diesem Modell sind in Abbildung 5-4 dargestellt.

Abbildung 5-4: German Data Trust



Quelle: Lind and Suckfüll (2013).

Der German Data Trust sieht die Festlegung von Regeln für die Nutzung und Monetarisierung personenbezogener digitaler Daten vor. Die Data Trust Instanz berücksichtigt alle Interessen und Präferenzen der beteiligten Akteure und strebt ein ausgewogenes Verhältnis der Eigentums- und Nutzungsrechte an, dem alle Wirtschaftsakteure zustimmen. Der Data Trust (als Treuhänder) verwaltet auch die von den Datenverarbeitern erhaltenen Nutzungsgebühren und stellt sicher, dass diese an die betroffenen Personen weitergegeben werden. Ein Vorteil dieses Ansatzes besteht darin, dass er die Kontrolle über die personenbezogenen Daten ermöglicht, die in einem verschlüsselten Format innerhalb der Daten-Trust-Einheit gespeichert sind, die per Definition keine wirtschaftlichen Interessen hat. Nach Heumann and Jentzsch (2019) kann ein solches Datenpooling ein Mittel zur Innovation sein. Aufgrund der regulatorischen Unsicherheit und fehlender Initiativen kleiner und mittlerer Unternehmen sind es bisher jedoch vor allem größere Firmen, die sich zu diesem Vorgehen entschieden haben.

5.3 Statische Datenportabilität und dynamische Interoperabilität

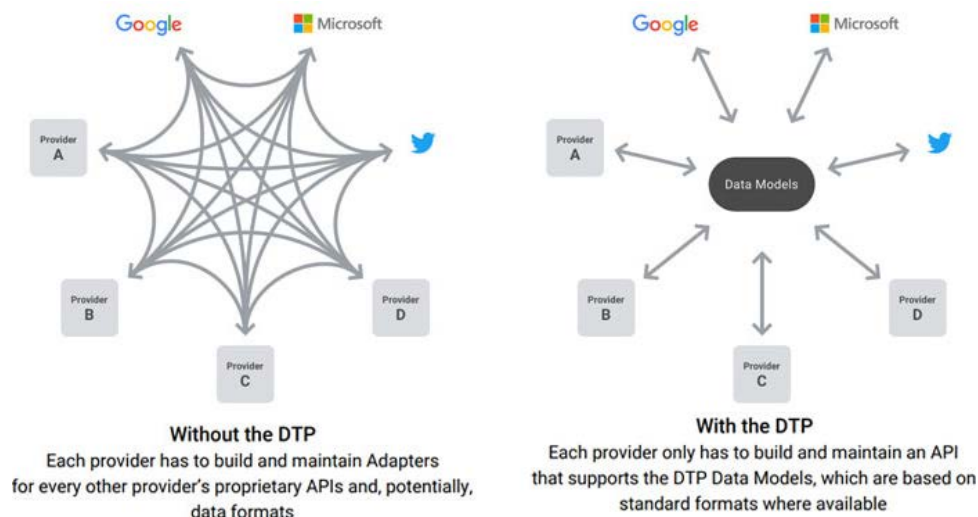
Erkenntnisse: Datenportabilität und Interoperabilität sind für die Innovation von zentraler Bedeutung. Während sich Datenportabilität meist auf offene Spezifikationen bezieht, ist Interoperabilität mit formalisierten Standards verbunden. Der Datenportabilität fehlen derzeit technische Standards bezüglich der Datenformate und Verarbeitungsprotokolle für die Datenextraktion und -implementierung, so dass sie derzeit kaum umgesetzt werden kann. Interoperabilitätsstandards können dazu beitragen, die Macht der Netzwerkeffekte und damit die Marktmacht der etablierten Unternehmen zu verringern, allerdings auf Kosten der wirtschaftlichen Effizienz. Sie können auch zu einem geringeren Maß an Vertraulichkeit und Datenschutz führen.

Das Recht auf **Datenübertragbarkeit** wird gemäß Artikel 20 der EU-Datenschutzgrundverordnung definiert als das Recht einer Person, „*die sie betreffenden personenbezogenen Daten, die sie einem Verantwortlichen bereitgestellt hat, in einem strukturierten, gängigen und maschinenlesbaren Format zu erhalten*“ und „*diese Daten einem anderen Verantwortlichen ohne Behinderung durch den Verantwortlichen, dem die personenbezogenen Daten zur bereitgestellt wurden, zu übermitteln*“. Folglich besteht der Prozess der Datenportabilität aus zwei Stufen: 1) Extraktion des Personendatensatzes aus dem Dienst A; und 2) Implementierung des Personendatensatzes in den Dienst B. Während der erste Teil (Datenextraktion) mit den aktuellen Datenschutzbestimmungen in Europa gut funktioniert, stellt sich das Hauptproblem bei der Implementierung des Datensatzes in einen anderen (Plattform-)Dienst. Heute gibt es keine technischen Standards, die von jedem Diensteanbieter verlangen, die gleichen, einheitlich angewandten Datenformate und Verarbeitungsprotokolle zu verwenden. Während sich Experten darüber einig sind, dass die Datenportabilität den Verbrauchern die Kontrolle über ihre Daten erleichtert, fehlt es derzeit an den technischen Möglichkeiten, um sie in der Praxis wirksam umzusetzen (Egan 2019). Daher kann ein Gesetz, das die Möglichkeit von Datentransfers vorschreibt, nur als ein erster Schritt zu echter Datenportabilität angesehen werden.

In dem Bemühen, diese Probleme zu überwinden, haben Apple, Deezer, Facebook, Google, Mastodon, Microsoft, Solid und Twitter das "Data Transfer Project (DTP)" ⁷² initiiert. Die Organisation arbeitet an Standards und Technologien, die eine reibungslose Realisierung der Datenportabilität weltweit gewährleisten. So verpflichten sich alle Teilnehmer des DTP, die gleichen Datenmodelle zu verwenden, damit Kunden bei Bedarf ihre personenbezogenen Daten nahezu in Echtzeit von Plattform A nach B übertragen können.

72 Siehe <https://datatransferproject.dev/>.

Abbildung 5-5: Das Data Transfer Project



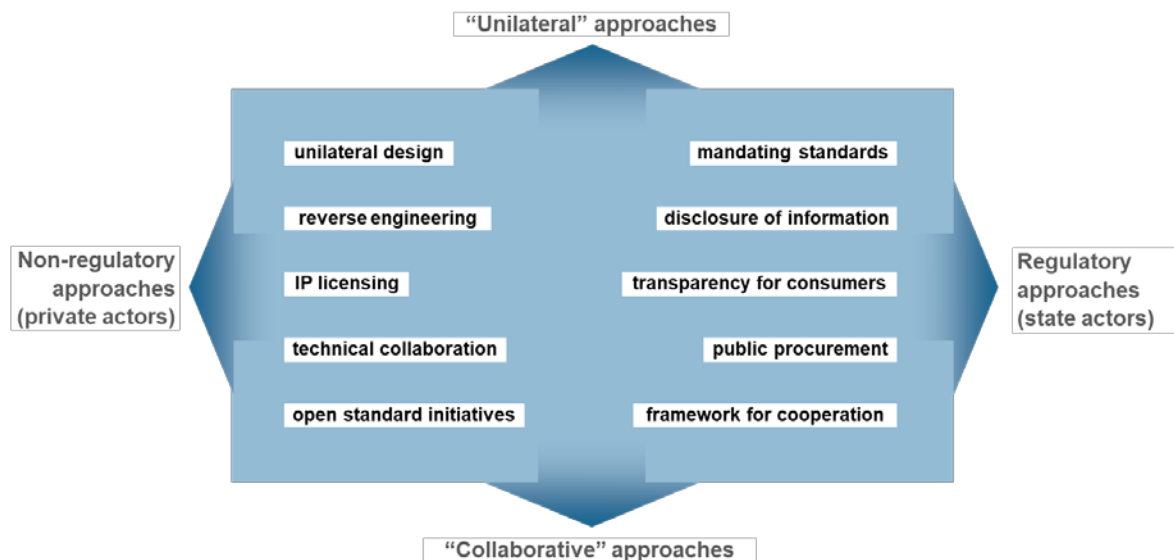
Quelle: <https://datatransferproject.dev/>

Eine solche Datenportabilität macht den Wechsel des Dienstleistungsanbieters durch die Verbraucher komfortabler und kann den Markteintritt neuer Unternehmen erleichtern. Die Datenschutzgrundverordnung gilt jedoch nur für Daten, die der Verbraucher (das Datensubjekt) "zur Verfügung stellt", wie beispielsweise Kaufdaten. Daten, die mit Hilfe von Datenanalyseverfahren von einem Unternehmen (dem für die Datenverarbeitung Verantwortlichen) "abgeleitet" werden, wie z.B. aus Kaufdaten hergeleitete Empfehlungen, fallen nicht unter die Datenschutzgrundverordnung. Lam and Liu (2018) zeigen, dass ohne Datenanalyse die Datenportabilität tatsächlich den Wechsel erleichtern kann, aber mit Datenanalytik kann die Datenportabilität den Wechsel behindern, weil die Verbraucher eine leichtere Datenportabilität erwarten und daher eher bereit sind, dem etablierten Betreiber Daten zur Verfügung zu stellen, was den Vorteil des etablierten Betreibers verstärkt. Swire and Lagos (2013) argumentieren, dass das Recht auf Datenportabilität den Software- und App-Anbietern erhebliche Kosten verursacht. Darüber hinaus könnten die Risiken der Datenportierbarkeit zu hoch sein, da ein einziger Identitätsmissbrauch zu einer lebenslangen Verletzung personenbezogener Daten führen kann. Wohlfarth (2019) zeigt, dass die Datenportabilität zwar zum Schutz der Nutzerinnen und Nutzer gedacht ist, dass sie jedoch Nachteile haben kann, weil Marktteilnehmer einen Anreiz haben, die Menge der gesammelten Daten im Vergleich zu einem System ohne Datenübertragbarkeit zu erhöhen. Die Umsätze für neue Dienste und der Gesamtgewinn steigen jedoch, wenn die Kosten für die Umsetzung nicht zu groß sind. Dies dürfte die Innovation und die Angebotsvielfalt verbessern.

Während sich die Datenportabilität meist auf offene Spezifikationen bezieht, ist die **Interoperabilität** mit formalisierten Standards verbunden. Interoperabilität kann definiert

werden als die "Fähigkeit, Daten und andere Informationen über Systeme, Anwendungen oder Komponenten hinweg zu übertragen und nutzbar zu machen" (Gasser & Palfrey 2007). Es gibt jedoch mehrere Ansätze für Interoperabilität, wie Abbildung 5-6 zeigt.

Abbildung 5-6: Lösungsansätze für die Interoperabilität



Quelle: Gasser and Palfrey (2007).

Die Arten der Interoperabilität unterscheiden sich nach Plattform, Gerät, Dienst, Datenformat, (Verarbeitungs-)Protokoll und Funktionalität. Je nach Art der Interoperabilität kann es vielfältige Auswirkungen auf Wettbewerb, Innovation und Datenschutz geben. Nach Casadesus-Masanell and Ruiz-Aliseda (2009) verändert beispielsweise Inkompatibilität das Kräfteverhältnis, und eine marktbeherrschende Plattform kann die Konkurrenz stärker überbieten, als dies bei Kompatibilität (d.h. Interoperabilität) der Fall wäre. Interoperabilität kann daher dazu beitragen, die wirtschaftliche Macht der Netzwerkeffekte und die Marktmacht der etablierten Unternehmen zu schwächen (Adner et al 2016). Gleichzeitig verringert sich dadurch jedoch auch die Wirtschaftlichkeit und kann je nach Ausgestaltung der Interoperabilitätsstandards auch zu einem geringeren Niveau an Vertraulichkeit und Datenschutz führen.

Es ist daher sehr wichtig, genau zu definieren, welche Art von Interoperabilität für das jeweilige Ziel in Frage kommt. Interoperabilität kann auf verschiedenen Ebenen relevant sein (Horak 2008): Syntaktische Interoperabilität bezieht sich auf die Möglichkeit, dass Systeme physisch miteinander verbunden werden und Daten austauschen können, während sich semantische Interoperabilität auf die Fähigkeit von Systemen bezieht, die Bedeutung der ausgetauschten Informationen zu verstehen. Horizontale Interoperabili-

tät (d.h. konkurrierende Dienste wie z.B. Mobilfunkdienste) und vertikale Interoperabilität (d.h. komplementäre Dienste wie z.B. ein Webbrowser und ein Betriebssystem) sind grundsätzlich möglich, wenn Daten/Informationen gemeinsam genutzt und/oder von komplementären bzw. konkurrierenden Diensten abgerufen werden können. In der Praxis existiert also ein Kontinuum zwischen voller und nicht vorhandener Interoperabilität, abhängig von der Anzahl der angegebenen Funktionalitäten.

In der Datenökonomie kann die Interoperabilität einer Vielzahl von Plattformen und Netzwerken der Schlüssel sein, insbesondere für Daten-Wertschöpfungsnetzwerke im Kontext von Industrie 4.0. Solche Marktinterventionen sind jedoch mit Nutzen und Kosten verbunden, und Interoperabilität und offene Systeme können für verschiedene Marktteilnehmer auch symmetrisch oder asymmetrisch sein. Sie müssen deshalb so gestaltet werden, dass eine Verhältnismäßigkeit zwischen Wettbewerb, Innovation und Datenschutz erzielt wird.

Die wichtigsten Ergebnisse von Kapitel 6

- *Die zentrale Herausforderung für die Datenökonomie scheint nicht der Datenzugriff als solcher zu sein, sondern die Fähigkeit, Datensätze Dritter zu finden, zu verstehen und zu integrieren.*
- *Die politischen Entscheidungsträger sollten die Entwicklung und Einführung gemeinsamer Datenarchitekturen als Mittel zur Vereinfachung des Datenaustauschs fördern.*
- *Referenzarchitekturen – sofern sie allgemein angenommen werden – werden voraussichtlich die Quantität und Qualität des Datenaustauschs erhöhen.*
- *Unternehmen sollten bevorzugt Daten von Drittanbietern verwenden, da sie im Allgemeinen billiger sind als die eigene Datenerfassung.*
- *Die politischen Entscheidungsträger sollten nicht darüber entscheiden, welche Art von Datenaustausch einen wirtschaftlichen Wert schaffen (Vermeiden der Entscheidung über Gewinner und Verlierer). Der Datenaustausch sollte marktgesteuert sein.*

6 Implikationen für die Wirtschaftspolitik und Regulierung

Bisher haben wir in der Debatte um die Datenökonomie eine verstärkte Aufmerksamkeit für Details gefordert. Im Gegensatz zu anderen Assets sind Daten äußerst heterogen, da sie innerhalb eines bestimmten Kontexts für einen bestimmten Zweck mit Hilfe eines ausgewählten Sensors oder einer Programmierschnittstelle (API) erfasst werden. Während sie dann analysiert werden, um wirtschaftlich verwertbare Informationen zu gewinnen, werden die Daten weiter verarbeitet und können als gehobene Daten in weitere Analysezyklen einfließen. Der zirkuläre Charakter der Wertschöpfung in der Datenökonomie und die vielfältigen Interaktionen zwischen den (oft mehrseitigen) Unternehmen führen zu weiterer Intransparenz.

Es ist klar, dass Daten eine wesentliche Rolle für Innovation und Wettbewerb spielen. Ob (und welche) Daten letztlich einen Wettbewerbsvorteil erzeugen können, hängt vom jeweiligen Anwendungszweck ab. Mit der zunehmenden Digitalisierung und den verschiedenen Möglichkeiten für Organisationen, Daten auszutauschen und wiederzuverwenden, ist eine grundsätzliche Hürde für den Datenzugang schwer vorstellbar. Aus unseren Ergebnissen geht jedoch auch hervor, dass mangelndes Wissen über die Einzelheiten der Datenerfassung, den Kontext und den Zweck, die (Vor-)Verarbeitung und andere Merkmale die Wiederverwendung von Daten für Unternehmen, Forscher und Behördenvertreter gleichermaßen einschränken. Eine größere Transparenz bezüglich dieser Merkmale würde wahrscheinlich den Datenfluss innerhalb und zwischen den Sektoren erhöhen. Im Gegenzug könnten die Beteiligten von einer größeren Menge wiederverwendbarer Daten sowie von einer besseren Effizienz der Wiederverwendung profitieren.

Während eine Senkung der Hürden für den Datenaustausch nicht notwendigerweise den Wettbewerb in der Datenökonomie stärken würde (da dieser von den jeweiligen Daten selbst und den Marktbedingungen abhängt), können regulatorische Verpflichtungen für Portierungsplattformen in Bezug auf personenbezogene Daten dazu beitragen, die Probleme bei der Datenimplementierung zu überwinden, die derzeit die Datenportabilität praktisch unmöglich machen. Daher wäre eine Standardisierung der Datenformate notwendig, um die Datenökonomie zu stärken. Allerdings ist keine Art von Datenaustausch ein Nullsummenspiel. Dort, wo sich aus der gemeinsamen Nutzung von Daten Vorteile ergeben, werden sich gewinnorientierte Unternehmen an solchen Aktivitäten beteiligen. Dagegen würde eine Regelung zur gemeinsamen Nutzung von Daten nicht die erforderliche Komplexität aufweisen und eine Entscheidung von Fall zu Fall erfordern. Marktgesteuerte bilaterale und multilaterale Vereinbarungen zwischen Unternehmen sind deshalb jeder allgemeinen (sektorspezifischen) Regulierung überlegen.

Eine allgemeine Regulierung zum Datenaustausch würde auch nicht das Datenschutzproblem lösen. Das Grundproblem des europäischen Datenschutzes, das es zu lösen gilt, ist das Instrument der Einwilligungserklärung selbst.

Die komplexe Struktur der Datenökonomie sollte politischen Entscheidungsträger Vorbehalte gegenüber einem allzu interventionistischen Ansatz vermitteln. Es besteht ein erhebliches Risiko, dass politische Entscheidungsträger (unbeabsichtigt) Gewinner und Verlierer auswählen. Angesichts der in diesem Bericht dargestellten Erkenntnisse erscheint ein Rahmen angemessen, der die zentrale Herausforderung der Transparenz in der Datenökonomie in einer Weise bewältigen kann, die dem grundlegenden politischen Prinzip der Technologieneutralität entspricht. Deshalb beginnt dieses Kapitel mit einer kurzen Zusammenfassung der offensichtlichen Herausforderungen in der Datenökonomie und schlägt weiterhin einen technologieneutralen politischen Ansatz zur Förderung einer europäischen Datenökonomie vor.

6.1 Politisch relevante Herausforderungen der Datenökonomie

***Erkenntnisse:** Die Datenzugänglichkeit an sich ist nicht die größte Hürde für eine erfolgreiche Datenökonomie. Transparenz, Auffindbarkeit und Wiederverwendbarkeit von Daten sind die wichtigsten Herausforderungen, die es zu bewältigen gilt. Interventionen müssen technologieneutral sein und allen Beteiligten den Datenaustausch ermöglichen.*

Dieser Bericht beleuchtet die wichtigsten Herausforderungen für eine erfolgreiche Datenökonomie. Wie in Kapitel 2 aufgezeigt wurde, mangelt es in der öffentlichen Debatte an einem gemeinsamen Verständnis dessen, was Daten ausmacht. Dies verhindert nicht nur ein differenziertes Verständnis des wirtschaftlichen Potenzials der Datenökonomie, sondern stellt auch den Austausch von Daten und ihre anschließende Wiederverwendung vor Herausforderungen. Dies ist in der Debatte um die Wiederverwendung von Forschungsdaten bereits deutlich geworden. Wie Borgman (2010) aufzeigt, erschwert die beträchtliche Bandbreite des Begriffs "Daten" unter den Kooperationspartnern und mehr noch zwischen den Disziplinen die Wiederverwendung von Forschungsdaten. Ebenso können Branchensilos und sogar Übergänge zwischen den einzelnen IT-Systemen das Datenaustauschpotenzial blockieren.

Daten von einem Kontext in einen anderen Kontext zu übertragen und sie dann im neuen Kontext nutzbar zu machen, dürfte erhebliche Investitionen in die Verarbeitung, Strukturierung oder Kennzeichnung der Daten in einem kompatiblen Format erfordern. Dabei besteht immer die Gefahr, dass die Daten nicht genutzt werden können, weil letztlich die spezifischen Eigenschaften der Daten, die sich aus der Aufbereitung während der Erfassung ergeben, nicht bekannt sind. In der Tat ist die Integration von Datensätzen aus verschiedenen Quellen eine zentrale Herausforderung für eine erfolgreiche Datenökonomie und insbesondere für die Innovation (Paunov & Planes-Satorra 2019).

Diese Investitionen, kombiniert mit den nicht unerheblichen Kosten für den Zugang zu relevanten Daten Dritter, können Unternehmen dazu veranlassen, bei der Erfassung der benötigten Daten auf ihre eigenen Kapazitäten zu vertrauen. Die Kontrolle über die eigene Datenerfassung reduziert die Kosten, die mit der Datenkonvertierung verbunden sind, sowie die Risiken, die sich aus einer mangelnden Kenntnis des Datenerfassungsprozesses ergeben. In datenintensiven Umgebungen dürften Unternehmen ihre eigenen Daten den Daten Dritter vorziehen, wann immer dies möglich ist. In datengerigen Umgebungen, in denen die Datenerfassung viel teurer ist als in datenreichen, sind die Effizienzgewinne durch Datenaustausch und -wiederverwendung insgesamt wesentlich größer. Da einige der wichtigsten Industriesektoren in Europa in solchen datenarmen Umgebungen tätig sind, verspricht die Schaffung von Rahmenbedingungen, die den Datenaustausch erleichtern können, erhebliche wirtschaftliche Effekte, z.B. durch die Freisetzung von Ressourcen zur Sammlung zusätzlicher Daten, anstatt von anderen gesammelte Daten zu reproduzieren oder sogar ihre Fehler bei der Anwendung der Daten zu vergrößern.⁷³ Andererseits kann insbesondere in datenarmen Umgebungen die Schaffung von Datenzugang an sich schon als eine Innovation betrachtet werden, die zu einem Wettbewerbsvorteil führt.

Vor diesem Hintergrund sollten die politischen Entscheidungsträger den Datenaustausch erleichtern, indem sie sicherstellen, dass sektorübergreifend Referenzarchitekturen und -standards entwickelt und angenommen werden. Bei diesen Bemühungen sollte der Kompromiss zwischen Transparenz und Datenzugang ebenso berücksichtigt werden wie die Kosten und die Komplexität der Möglichkeit, wichtige Geschäftsgeheimnisse der beteiligten Organisationen zu wahren. Insbesondere wird die Komplexität des Schutzes von Betriebsgeheimnissen mit fortschreitender Digitalisierung zunehmen.

Wenn Daten in noch größerem Umfang als heute ausgetauscht werden, können sich Organisationen auch dazu veranlasst sehen, mehr in eine qualitativ hochwertige Datenerfassung und umfassendere Metadaten zu investieren, da dies den Wert ihrer Daten für den späteren Austausch erhöht. Ein Rechtsrahmen zur Erleichterung des Datenaustauschs kann daher auch die Verfügbarkeit von qualitativ hochwertigen Daten verbessern. Da jedoch immer mehr und vielfältigere Daten in großem Umfang verfügbar werden, wird das Auffinden der jeweils geeigneten Daten eine immer wichtigere Rolle für eine erfolgreiche Datenökonomie spielen. Jede politische Maßnahme zur Förderung der europäischen Datenökonomie sollte einen Rahmen vorgeben, der eine effiziente Datenfindung ermöglicht.

Schließlich benötigen Unternehmen Zugang zu Talenten mit den für die Datenanalytik erforderlichen Fähigkeiten. Der offensichtliche Mangel an diesen Fähigkeiten kann die Entwicklung der europäischen Datenökonomie bremsen.

⁷³ Dies ist besonders dann relevant, wenn die Innovation auf Daten beruht, deren Erlangung kostspielig ist, z.B. bei pharmazeutischen Studien. Siehe z.B. Federico G, Morton FS, Shapiro C. 2019. Antitrust and Innovation: Welcoming and Protecting Disruption. In *Innovation Policy and the Economy, Volume 20*: University of Chicago Press.

Boyd and Crawford (2012) weisen darauf hin, dass es an den Universitäten einen besonders ausgeprägten Mangel an Erfahrungen mit Big Data gibt, vor allem wenn sich nur Eliteuniversitäten den Zugang zu Big Data leisten können.

Aufgrund der in diesem Bericht gesammelten Erkenntnisse sind wir der Meinung, dass die Förderung von Referenzarchitekturen viele der skizzierten Herausforderungen angehen kann. Im folgenden Abschnitt wird der Vorschlag im Detail erörtert, wobei relevante Beispiele für solche Referenzarchitekturen vorgestellt werden.

6.2 Referenzarchitekturen als Weg in die Zukunft

Erkenntnisse: *Referenzarchitekturen bieten einen technologieneutralen Rahmen, der die wichtigsten Probleme für einen häufigen und qualitativ hochwertigen Datenaustausch beheben kann. Die politischen Entscheidungsträger sollten die Entwicklung solcher Architekturen sektorübergreifend fördern, um die europäische Datenökonomie zu fördern.*

Auf der Grundlage der in dieser Studie gewonnenen Erkenntnisse und insbesondere der im vorangegangenen Abschnitt skizzierten zentralen Herausforderungen für eine Datenökonomie schlagen wir die Förderung von Referenzarchitekturen vor, um die geeigneten Rahmenbedingungen für häufigeren und besseren Datenaustausch innerhalb und zwischen den Sektoren zu schaffen.

Solche Referenzarchitekturen ermöglichen im Wesentlichen eine konsistente Darstellung der komplexen Beziehungen innerhalb von Wertschöpfungsnetzwerken, die in der Datenökonomie entstehen, wie in Abschnitt 3.1 beschrieben. Das heißt, eine Referenzarchitektur besteht aus mehreren Schichten und ist eine gemeinsame Sprache für Daten, Prozesse und Schnittstellen, die über Abteilungs-, Unternehmens-, Branchen- und Sektorengrenzen hinausgeht. Eine solche gemeinsame Sprache hilft, Transaktionskosten zu sparen und ermöglicht eine höhere Effizienz als die heute üblichen fragmentierten IT-Infrastrukturen. Insbesondere der in Abschnitt 5.2 beschriebene industrielle Datenraum implementiert diese Logik bereits in einer vertikalen Datenpooling-Umgebung.

Die Notwendigkeit einer gemeinsamen Sprache für den Datenaustausch liegt auf der Hand: Srai et al (2019) argumentieren, dass "*Lieferketten-Ontologien zwar leicht zu übersehen sind, aber das Rückgrat einer implementierbaren [digitalen Zwillinge]-Lieferkette bilden*" (S. 3). Cuquet and Fensel (2018) sind der Ansicht, dass die Datensemantik als Forschungs- und Innovationsthematik einen zentralen Baustein darstellen sollte, um in Zukunft große und positive Auswirkungen von Big Data in Europa zu erzielen. Die Datensemantik spielt eine entscheidende Rolle für verbesserte Effizienz, Innovation, sich ändernde Geschäftsmodelle, Beschäftigung, öffentliche Finanzierung sowie Bewusstseinsbildung. Im Großen und Ganzen sind sich die Experten aus Wissenschaft und Industrie einig, dass Kompatibilitäts- und Interoperabilitätsstandards erforderlich sind, um aus Daten einen ökonomischen Wert zu schaffen und eine europäische Datenökonomie auf breiterer Basis zu stärken. Es ist wichtig, dass die Akteure über die Fähigkeit und das Potenzial verfügen, relevante Datensätze zu verknüpfen und zu aggregieren. Dies erfordert die Entwicklung und Umsetzung einheitlicher Standards, die die Interoperabilität von Daten ermöglichen (HM Treasury 2018). Standards als solche sind jedoch möglicherweise nicht ausreichend.

Ein ganzheitlicher Rahmen in Form einer Referenzarchitektur ist möglicherweise angemessener.⁷⁴ Referenzarchitekturen können den Datentransfer von einem Kontext in einen anderen ermöglichen. Da sie relevante Metadaten in standardisierten Formaten enthalten, können die beteiligten Organisationen diese Informationen nutzen, um zu verstehen, wie die Daten erfasst wurden, und können sie gegebenenfalls bis zu ihrer Quelle zurückverfolgen. Einmal eingerichtet, können Referenzarchitekturen die Kosten für die Erstellung solcher Metadaten aufgrund ihres dann standardisierten Charakters erheblich reduzieren. Wenn sie Informationen über die Lizenzierung der ausgetauschten Daten enthalten, könnten sie auch den Bedenken Rechnung tragen, die hinsichtlich des Schutzes relevanter Geschäftsgeheimnisse aufkommen.⁷⁵ Carrière-Swallow and Haksar (2019) schlagen beispielsweise Lizenzierungssysteme vor, die diese Bedenken mildern könnten, da sie die Zwecke, für die die Daten verwendet werden dürfen, eindeutig festlegen könnten.⁷⁶ Sie können auch die Auffindbarkeit potenzieller Mitarbeiter und ihrer Fähigkeiten sowie die Entdeckung von Datenangeboten verbessern, da eine Hauptfunktion einer Referenzarchitektur darin besteht, eine logische Positionierung und Klassifizierung neuer Technologien, Standards und Akteure innerhalb von Wertschöpfungsnetzwerken bereitzustellen.

Referenzarchitekturen sind daher ein entscheidender Schritt zur Realisierung der Vision von Industrie 4.0. Sie sind in der Lage, die IT-Systeme, die technischen Ressourcen sowie die Fähigkeiten und Kapazitäten verschiedener Unternehmen und Branchen innerhalb eines Wertschöpfungsnetzwerks zu koordinieren. Eine Referenzarchitektur hat die Aufgabe, einen Rahmen für die Strukturierung, Entwicklung und Implementierung eines Wertschöpfungsnetzwerks zu schaffen und kann auch die Integration sowie den Betrieb der Informationssysteme unterstützen. Eine übergreifende Referenzarchitektur – die die Grenzen zwischen industriellen Wertschöpfungsketten und schließlich Industriesektoren aufhebt – kann ein Weg in die Zukunft sein.

Um die Auswirkungen von Referenzarchitekturmodellen zu verstehen, können die politischen Entscheidungsträger auf den Automobilsektor blicken, der zu den ersten Sektoren gehörte, die eine solche Architektur eingeführt haben.⁷⁷ Die Referenzarchitektur ist als "AUTOSAR" bekannt. Sie begann 2003 als eine Partnerschaft bedeutender Akteure im Automobilsektor, die an der Entwicklung offener Industriestandards für eine elektri-

⁷⁴ Insbesondere kann ein solcher Rahmen nicht jede mögliche zukünftige Wiederverwendung von Daten vorhersehen. Ein optimaler Rahmen ist genauso schwer zu definieren wie eine optimale Sprache. Es wird immer Mehrdeutigkeiten geben: Laporte S. 2018. *Ideal language*. *KO KNOWLEDGE ORGANIZATION* 45: 586-608. Dennoch sind wesentliche Verbesserungen für den Datenaustausch und die Wertschöpfung aus Daten zu erwarten, wie die folgenden Beispiele im Haupttext zeigen.

⁷⁵ Intelligente Verträge, die z.B. auf der Blockchaintechologie basieren, könnten in einem solchen Rahmen, wie er von Roman D., Stefano G. 2016 skizziert wurde, eine wichtige Rolle spielen. Roman D, Stefano G. 2016. *Towards a Reference Architecture for Trusted Data Marketplaces: The Credit Scoring Perspective*. Presented at 2nd International Conference on Open and Big Data (OBD), Vienna.

⁷⁶ Insbesondere die Durchsetzung solcher Systeme könnte in der Praxis schwierig sein, da der Datenmissbrauch nicht-rival und praktisch nicht rückverfolgbar ist. Eine florierende Data Economy muss daher auf dem Vertrauen zwischen den beteiligten Akteuren aufbauen.

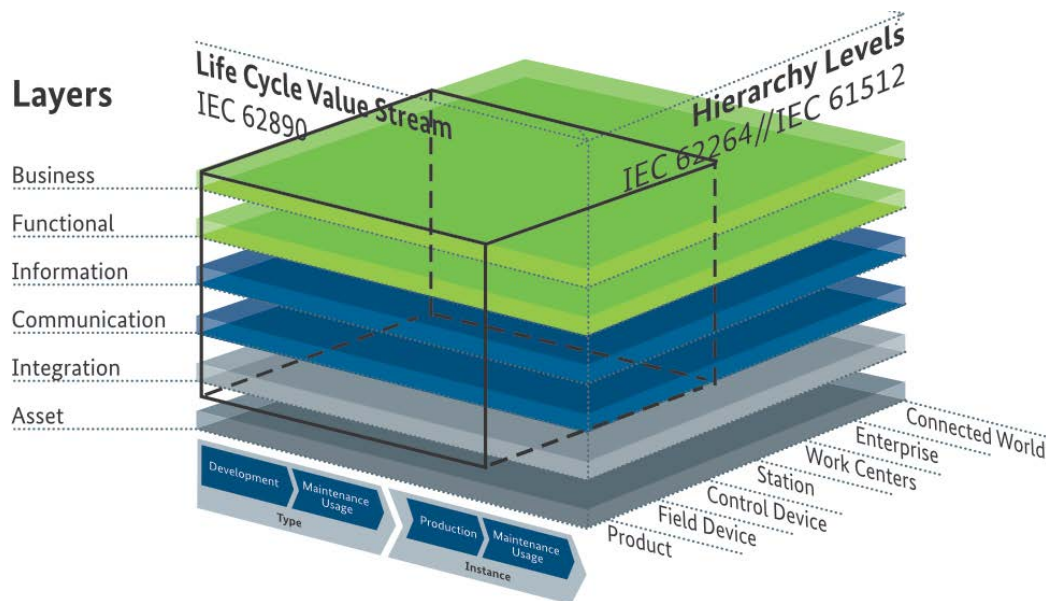
⁷⁷ Industrielle Referenzarchitekturen wurden auch in der Landwirtschaft "ISOBUS", für intelligente Stromnetze und im Gesundheitssektor eingerichtet.

sche/elektronische (E/E) Architektur interessiert waren. In der Anfangsphase waren die BMW Group, Bosch, Continental, DaimlerChrysler, Ford Motor Company, PSA, Siemens VDO, Toyota Motor Company und Volkswagen Mitglieder der Partnerschaft (Heinecke et al 2004). Sie hatten erkannt, dass die automobilen E/E-Systeme immer wieder neue Funktionen hinzufügen würden und dass dies wiederum die Komplexität und die Kosten für jeden einzelnen von ihnen exponentiell erhöhen würde, wenn sie an der Vielzahl der damals vorherrschenden proprietären Standards festhielten. Es dauerte bis 2005 bis zur ersten Veröffentlichung der (inzwischen klassischen) AUTOSAR-Plattform.⁷⁸ Zu Beginn der Phase 2 waren 166 Partner AUTOSAR beigetreten, im Dezember 2016 waren es 191 Partner. Die Plattform wurde zu einer Referenzarchitektur weiterentwickelt und wird derzeit überarbeitet und zu einer adaptiven Plattform ausgebaut.⁷⁹ Ein Literaturüberblick von Dersten et al (2011) zeigt sowohl positive als auch negative Auswirkungen in verschiedenen Bereichen auf. So hat beispielsweise die Implementierung von AUTOSAR die Komplexität reduziert, ein verbessertes Qualitätsmanagement ermöglicht und die Zuverlässigkeit der Software erhöht. Sie verbesserte auch die Position der Zulieferer, die leichter Lösungen für verschiedene OEMs entwickeln konnten. AUTOSAR führte jedoch auch zu Performance-Risiken und einem Trade-off zwischen Speicher- und Reaktionszeiten in E/E-Systemen. Innerhalb der Unternehmen mussten neue Prozesse entwickelt werden, die erhebliche Investitionen und interne Reibungsverluste mit sich brachten. Der anhaltende Erfolg einer ganzheitlichen Architektur im Automobilssektor wird jedoch durch die jüngsten Analysen von Continental unterstrichen, die behaupten, dass fortschrittliche Architekturen die Innovation im Automobilssektor unterstützen. Beispielsweise kann durch die Anwendung einer Fahrzeugsystemarchitektur etwa ein Drittel der Leitungen innerhalb eines modernen Fahrzeugs eingespart werden, was zu einer signifikanten Gewichtsreduzierung und damit zu einer Reduzierung des Kraftstoffverbrauchs und der Emissionen führt (Locks & Winkler 2017).

⁷⁸ Release 1.0.0 wurde am 26. Juni 2005 veröffentlicht.

⁷⁹ <https://www.autosar.org/about/history/>.

Abbildung 6-1: RAMI4.0 als Beispiel für eine Referenzarchitektur für Industrie 4.0



Quelle: DIN SPEC 913 cf. Arnold and Liebe (2018).

Während traditionelle Referenzarchitekturen etwas auf einzelne Sicherheitselemente beschränkt blieben, wurde in Deutschland im Rahmen der Plattform Industrie 4.0 ein übergreifendes Referenzarchitekturmodell für Industrie 4.0 (RAMI 4.0)⁸⁰ entwickelt.⁸¹ Abbildung 6-1 stellt das Schichtenmodell dar. Mit RAMI 4.0 wird schließlich auch kleinen und mittleren Unternehmen (KMU) ein Werkzeug an die Hand gegeben, das die Einführung und Umsetzung von herstellerunabhängigen IT-Lösungen und den Datenaustausch ermöglicht und fördert. Die Vereinbarung von Standards ist ein wesentlicher Aspekt des unternehmensübergreifenden Datenaustauschs, der wiederum die Kommunikation zwischen Maschinen oder zwischen Komponenten und Maschinen ermöglicht. Nur wenn es eine gemeinsame Semantik der Daten gibt, kann das Netzwerk auch horizontal funktionieren.⁸²

⁸⁰ Für ausführlichere Informationen zu RAMI 4.0 und seinen Möglichkeiten siehe Arnold R, Liebe A. 2018. Digitale Wertschöpfungsnetzwerke und RAMI 4.0 im hessischen Mittelstand, Hessisches Ministerium für Wirtschaft, Energie, Verkehr und Landesentwicklung, Wiesbaden.

⁸¹ Folgende Akteure waren beteiligt: Die Verbände BITKOM, VDMA, ZVEI und VDI kooperierten mit verschiedenen Forschungseinrichtungen und Geschäftspartnern unter der Koordination des Arbeitskreises "Referenzarchitekturen, Standards und Normung" der Plattform Industrie 4.0."

⁸² Ein wesentlicher Schritt in diese Richtung ist die Integration der Open Platform Communications Unified Architecture (OPC UA) in RAMI 4.0. Der OPC UA-Standard fügt sich nahtlos in RAMI 4.0 ein. Mit der sukzessiven Integration von Standards gewinnt RAMI 4.0 sowohl an Bedeutung als auch an Einfluss.

Wie bereits in Abschnitt 3.3 über Innovationen in der Datenökonomie erwähnt, steht das Konzept des digitalen Zwillings im Mittelpunkt der Vision von Industrie 4.0. Es bietet eine einheitliche digitale Repräsentation der Eigenschaften von Assets innerhalb eines komplexen Wertschöpfungsnetzwerks. Dies steht in deutlichem Kontrast zu den heutigen anwendungszentrierten Lösungen, die meist auf sehr heterogenen proprietären Systemen basieren. Wenn Daten über solche Systeme ausgetauscht werden, sind Medienbrüche sowie inkonsistente und redundante Daten üblich. Geringe Datenqualität und mangelnde Effizienz des Datenaustauschs sind die Folge. Das Konzept des digitalen Zwillings wird in RAMI 4.0 als Verwaltungsschale aufgegriffen. Die Verwaltungsschale ermöglicht die vollständige digitale Repräsentation von physischen Einheiten sowohl in der horizontalen Vernetzung, d.h. unternehmensübergreifend, als auch in der vertikalen Vernetzung, d.h. innerhalb einer Einheit.

Als Erweiterung bieten Referenzarchitekturen die Grundstruktur und den Rahmen für industrielle Datenplattformen (IDPs). Beispiele für IDPs sind die Cooperative ITS for Mobility in European Cities (CIMEC) und openEASE. IDPs bieten eine Möglichkeit, sektorrelevante Daten zusammenzuführen und sind von besonderem Wert für die Entwicklung neuer Dienste, Anwendungen der künstlichen Intelligenz (KI) und innovativer Geschäftsprozesse. Der Schlüsselfaktor für den Erfolg dieser Projekte besteht darin, ein Gleichgewicht zwischen der effektiven und ganzheitlichen gemeinsamen Nutzung von Daten und der Wahrung der digitalen Souveränität der Teilnehmer zu finden, die für viele Beteiligte eines der wesentlichen Hindernisse für die Teilnahme am Datenaustausch darstellt.

Die potenziellen Auswirkungen solcher Referenzarchitekturen gehen jedoch weit über die industrielle Anwendung hinaus, wie Beispiele aus Estland und Südkorea zeigen. Der Erfolg der estnischen eGovernment-Implementierung beruht zum Teil auf der dort verwendeten Referenzarchitektur. Bereits 2005 hat das estnische Wirtschaftsministerium ein Strategiepapier zur Umsetzung der semantischen Interoperabilität der staatlichen Informationstechnologie herausgegeben. Die Strategie konzentrierte sich auf Datenobjekte und Ein-/Ausgabeparameter von Datendiensten, die alle innerhalb des Systems semantisch beschrieben werden. Die Regierung unterstützt diese Bemühungen kontinuierlich mit entsprechenden Richtlinien, Vorgaben, Werkzeugen sowie Bildungs- und Werbeaktivitäten. Die Referenzarchitektur wurde schließlich im Jahr 2009 eingerichtet. Anfangs war der Erfolg relativ begrenzt. Eine wesentliche Hürde war die Komplexität der bestehenden Ontologien, mit der die Behörden nicht zurechtkamen; deshalb wurde eine spezifische Ontologie entwickelt. Bis 2011 wurden jährlich 240 Millionen Online-Datenanfragen von estnischen Bürgern und öffentlichen Diensten gestellt, die durch die Referenzarchitektur ermöglicht werden (Haav & Küngas 2013). Darunter wurden 94% der Steuererklärungen online abgegeben (cf Kütt & Priisalu 2014). Heute sind 99% der Dienstleistungen online.⁸³

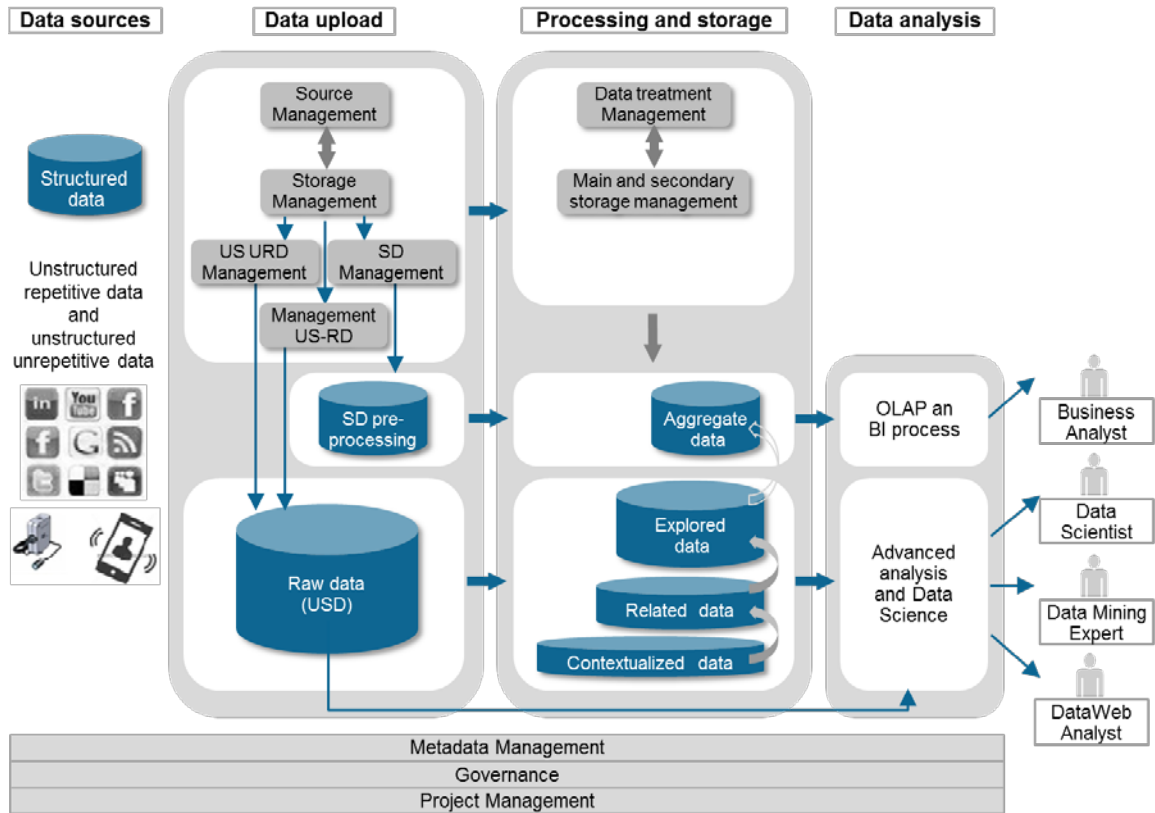
⁸³ <https://e-estonia.com/>

Das *Korean National Standard Reference Data (SRD) Program* ist etwas weniger umfassend als die in Estland entwickelte Architektur und kann nur als Teil einer Referenzarchitektur betrachtet werden. Dennoch verweist es auf eine potenzielle Best Practice der Regierung, die den Datenaustausch durch die Förderung der Standardisierung und die Schaffung eines Rahmens für einen sicheren Datenaustausch erleichtert. Das SRD wurde 1999 gesetzlich verankert und 2006 umgesetzt. Eine Schlüsselkomponente des Programms war die Festlegung eindeutiger Richtlinien zur Bewertung der Daten(qualität). In den letzten zehn Jahren wurden 43.000 Datenbanken entwickelt, die Teil der SRD für wissenschaftliche, soziale und industrielle Bereiche geworden sind (Lee 2019).

Die in diesem Bericht gesammelten Erkenntnisse sowie die Beispiele von Referenzarchitekturen und ähnlichen Datenaustauscharchitekturen, die einem ähnlichen Zweck dienen, verdeutlichen, dass ohne ihre Metadaten ausgetauschte Daten in der Tat viel weniger wertvoll sein können als klar in eine definierte Ontologie eingebettete Daten. Der Zugang zu großen Nutzer-Pools kann einige dieser Probleme mildern, da die Nutzer aufgefordert werden können, Daten mit Bezug auf bestimmte Inhalte oder andere Auswertungsmöglichkeiten zu "taggen". Man geht davon aus, dass sie letztendlich die richtigen Ergebnisse erzeugen, da die Mehrheit der Nutzer die richtige Interpretation der Daten liefert. Allerdings kann dieser Ansatz in einigen Kontexten gut funktionieren, in anderen jedoch nicht, da die Nutzer ihre kulturelle Voreingenommenheit in die Aufgabe einbringen könnten, was für den Zweck, für den die Daten verwendet werden sollen, nützlich sein kann oder auch nicht (Caracciolo et al 2018, Stuckenschmidt 2012). Ein weiterer Vorteil der Integration spezifischer Ontologien in Referenzarchitekturen besteht darin, dass die Metadaten maschinenlesbar und in jedem Informationssystem mit den Daten verknüpft bleiben können. Dies trifft auf die meisten Implementierungen bestehender Big Data-Lösungen, bei denen die Metadaten typischerweise in die Inhaltsdaten eingebettet sind, nicht zu (Marshall 2012).

Bei der Verfolgung eines solchen Ansatzes sollten die politischen Entscheidungsträger die bereits im Bereich des Data Warehousing gemachten Erfahrungen nutzen, wie sie beispielsweise im Hinblick auf die sich abzeichnenden Anforderungen an Big Data von Inmon (2006), Inmon et al (2010) oder Kimball (2011) aufgezeigt wurden. Unter Berücksichtigung dieser Erkenntnisse haben Salinas and Lemus (2017) ein mehrschichtiges, gestaffeltes Architekturmodell für Big Data vorgeschlagen (s. Abbildung 6-2). Das Modell berücksichtigt die Verarbeitung aller Arten von Daten: Erfassung, Bereinigung, Integration, Identifizierung, Analyse und Kontrolle der Datenqualität. Es umfasst auch transversale Komponenten für Datenspeicherung, Metadaten, Lebenszyklus und Sicherheitsmanagement.

Abbildung 6-2: Ein mehrschichtiges gestaffeltes Architekturmodell für Big Data



Quelle: Salinas and Lemus (2017).

7 Fazit

Diese Studie hatte zum Ziel, die Debatte über die Datenökonomie in Europa zu vertiefen, wobei ein besonderer Schwerpunkt auf der Untersuchung möglicher Bedenken hinsichtlich der Kontrolle über Daten und der Einführung eines möglicherweise obligatorischen Zugangs zu Daten lag. Es wurde festgestellt, dass es in der öffentlichen Diskussion an Differenziertheit mangelt und mitunter falsche Annahmen über die Datenökonomie vorherrschen. Insbesondere die großen Unterschiede zwischen Daten und die Rolle, die die Datenqualität spielt, werden zugunsten einer Fokussierung auf die Datenmenge vernachlässigt.

Um der Debatte eine fundierte Grundlage zu verleihen, haben wir ein mehrschichtiges Modell entworfen, das zeigt, wie Daten erfasst und verarbeitet und wie Erkenntnisse gewonnen werden, aus denen wirtschaftlich verwertbare Informationen generiert werden. Unser Modell hebt die Komplexität und die verschiedenen Interaktionen hervor, die Teil des Prozesses sind. Es unterstreicht ferner, dass Datenerhebung von Natur aus konstitutiv ist, so dass die Annahme, Daten seien objektive Fakten, nicht haltbar ist. Dies hat Auswirkungen auf den Wert der Daten als solche, da Daten in der Regel nur in einem bestimmten Kontext und für einen bestimmten Zweck, für den sie erfasst wurden, einen ökonomischen Wert generieren. Dies schließt zwar nicht aus, dass die Daten in einem anderen Kontext und für einen anderen Zweck verwendet werden, macht aber deutlich, dass es Kosten verursacht, sie dafür nutzbar zu machen.

Aufgrund der rasch zunehmenden Digitalisierung und der bestehenden Möglichkeiten des Datenaustauschs zwischen den Unternehmen, die Daten erheben, und Dritten (Unternehmen, denen der Zugang zu den Daten gewährt wird, die jedoch keine unmittelbare Kontrolle über den Prozess der Datenerfassung haben), scheint der Datenzugang als solcher kein Hindernis für die Datenökonomie im Allgemeinen zu sein. Die Förderung des Datenaustauschs hätte positive wirtschaftliche Auswirkungen, da die Unternehmen effizienter arbeiten und vielfältigere Daten erfasst werden könnten, da doppelte Arbeit vermieden werden könnte.

Ein interventionistischer Ansatz scheint jedoch nicht angemessen zu sein. Stattdessen empfehlen wir, die Entwicklung von Referenzarchitekturen zur Schaffung von Kompatibilität und Interoperabilität der IT-Infrastrukturen sowie Datenformaten und Verarbeitungsprotokollen aktiv zu fördern. Positive Beispiele aus der Industrie sowie von Regierungen (Datenmanagement-Architekturen in Estland und Südkorea) unterstreichen den potenziellen Erfolg solcher Rahmenwerke. Letztlich sollte die Entwicklung solcher Referenzarchitekturen natürlich bei den Marktakteuren und den Normungsorganisationen verbleiben.

Eine derartige Förderung des Engagements der Industrie müsste natürlich durch Bemühungen zur Vergrößerung des Talentpools und zur Förderung eines allgemein unterstützenden Geschäftsumfelds ergänzt werden, einschließlich ausreichender Risiko-

kapitalmittel, F&E-Unterstützung und einer angemessenen Regulierung des Datenschutzes und der Sicherheit. Hierzu ist eine konzertierte Anstrengung von Institutionen und Behörden auf der Ebene der Mitgliedstaaten und auf europäischer Ebene erforderlich.

Literatur

- Adner R, Chen J, Zhu F. 2016. Frenemies in Platform Markets: The Case of Apple's iPad vs. Amazon's Kindle
- Adomavicius G, Bockstedt JC, Curley SP, Zhang J. 2017. Effects of online recommendations on consumers' willingness to pay. *Information Systems Research* 29: 84-102
- Aguirre E, Mahr D, Grewal D, de Ruyter K, Wetzels M. 2015. Unraveling the Personalization Paradox: The Effect of Information Collection and Trust-Building Strategies on Online Advertisement Effectiveness. *Journal of Retailing* 91: 34-49
- Ansari A, Mela CF. 2003. E-customization. *Journal of Marketing Research*, 40: 131-45
- Anthes G. 2015. Data Brokers are Watching You. *Communications of the ACM* 58: 28-30
- Areeda P, Hovorkamp H. 1988. An Analysis of Antitrust Principles and Their Application. *Antitrust Law* 736: 675-701
- Argenton C, Prüfer J. 2012. Search engine competition with network externalities. *Journal of Competition Law & Economics* 8: 73-105
- Armstrong M. 2006. Competition in Two-Sided Markets. *The RAND Journal of Economics* 37: 668-91
- Arnaut C, Pont M, Scaria E, Berghmans A, Leconte S. 2018. Study on data sharing between companies in Europe. A study prepared for the European Commission, European Commission; everis, Luxembourg
- Arnold R, Bott J, Hildebrandt C, Schäfer S, Tenbrock S. 2016. Internet-basierte Plattformen und ihre Bedeutung in Deutschland, Wissenschaftliches Institut für Infrastruktur und Kommunikationsdienste (WIK), Bad Honnef
- Arnold R, Hildebrandt C. 2017. The Socio-Economic Impact of Online Platforms, Wissenschaftliches Institut für Infrastruktur und Kommunikationsdienste (WIK), Bad Honnef
- Arnold R, Liebe A. 2018. Digitale Wertschöpfungsnetzwerke und RAMI 4.0 im hessischen Mittelstand, Hessisches Ministerium für Wirtschaft, Energie, Verkehr und Landesentwicklung, Wiesbaden
- Arnold R, Schiffer M, Pols A. 2013. Wirtschaft Digitalisiert - Welche Rolle spielt das Internet für die deutsche Industrie und Dienstleister?, IW Consult and BITKOM, Cologne, Berlin
- Arnold R, Waldburger M. 2015. The Economic Influence of Data and their Impact on Business Models In *Trends in Telecommunication Reform 2015 - Getting Ready for the Digital Economy*, ed. ITU, pp. 153-83. Geneva: International Telecommunication Union
- Attard J, Orlandi F, Auer S. 2016 *IEEE/WIC/ACM International Conference on Web Intelligence (WI)*, 13-16 Oct. 2016: 453-56. IEEE.
- Attard J, Orlandi F, Auer S. *International Conference on Theory and Practice of Electronic Governance, New Delhi, India, 2017*: 475-784.
- Austin PC, Goldwasser MA. 2008. Pisces did not have Increased Heart Failure: Data-driven Comparisons of Binary Proportions between Levels of a Categorical Variable can Result in Incorrect Statistical Significance Levels. *Journal of Clinical Epidemiology* 61: 295-300
- Austin PC, Mamdani MM, Juurlink DN, Hux JE. 2006. Testing Multiple Statistical Hypotheses Resulted in Spurious Associations: a Study of Astrological Signs and Health. *Journal of Clinical Epidemiology* 59: 871-72
- Awad NF, Krishnan MS. 2006. The Personalization Privacy Paradox: An Empirical Evaluation of Information Transparency and the Willingness to Be Profiled Online for Personalization. *MIS Quarterly* 30: 13-28

- Banko M, Brill E. *Proceedings of the 39th Annual Meeting on Association for Computational Linguistics 2001*: 26-33. Association for Computational Linguistics.
- Barbero M, Bartz K, Linz F, Mauritz S, Wauters P, et al. 2018a. Study to support the review of Directive 2003/98/EC on the re-use of public sector information, Deloitte, Brussels
- Barbero M, Cocoru D, Graux H, Hillebrand A, Linz F, et al. 2018b. Study on emerging issues of data ownership, interoperability, (re-)usability and access to data, and liability, Deloitte, Brussels
- Barney J. 1991. Firm resources and sustained competitive advantage. *Journal of management* 17: 99-120
- Benbasat I, Wang W. 2005. Trust in and adoption of online recommendation agents. *Journal of the association for information systems* 6: 4
- Benlian A. 2015. Web personalization cues and their differential effects on user assessments of website value. *Journal of Management Information Systems* 32: 225-60
- BEREC. 2019. BEREC Report on the Datenökonomie, Body of European Regulators for Electronic Communications
- Berente N, Seidel S, Safadi H. 2018. Research Commentary—Data-Driven Computationally Intensive Theory Development. *Information Systems Research* 30: 50-64
- Bergemann D, Bonatti A. 2019. Markets for Information: An Introduction. *Annual Review of Economics* 11: 85-107
- Bernstein ES. 2012. The transparency paradox: A role for privacy in organizational learning and operational control. *Administrative Science Quarterly* 57: 181-216
- Bertenrath R, Arnold R, Koppel O, Lang T. 2011. Innovation Policy and the Business Cycle: Innovation Policy's Role in Addressing Economic Downturn - INNO-Grips Policy Brief No. 1, European Commission, Cologne/Brussels
- Blank G, Lutz C. 2017. Representativeness of social media in Great Britain: investigating Facebook, LinkedIn, Twitter, Pinterest, Google+, and Instagram. *American Behavioral Scientist* 61: 741-56
- Bleier A, Eisenbeiss M. 2015a. The importance of trust for personalized online advertising. *Journal of Retailing* 91: 390-409
- Bleier A, Eisenbeiss M. 2015b. Personalized online advertising effectiveness: The interplay of what, when, and where. *Marketing Science* 34: 669-88
- Boerman SC, Kruikemeier S, Zuiderveen Borgesius FJ. 2017. Online Behavioral Advertising: A Literature Review and Research Agenda. *Journal of Advertising* 46: 363-76
- Borgman CL. 2010. Research Data: Who will share what, with whom, when, and why? - (RatSWD Working Paper Series Nr. 161, Berlin
- Bott J, Hildebrandt C, Arnold R. 2018. Die Nutzung von Daten durch OTT-Dienste zur Abschöpfung von Aufmerksamkeit und Zahlungsbereitschaft: Implikationen für Wettbewerb, Regulierung sowie Daten- und Verbraucherschutz - WIK-Diskussionsbeitrag Nr. 431, Wissenschaftliches Institut für Infrastruktur und Kommunikationsdienste (WIK), Bad Honnef
- Boyd D, Crawford K. 2012. Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon. *Information, Communication & Society* 15: 662-79
- Brandimarte L, Acquisti A, Loewenstein G. 2012. Misplaced Confidences: Privacy and the Control Paradox. *Social Psychological and Personality Science* 4: 340-47
- Bründl S, Matt C, Hess T. 2015. Wertschöpfung in Datenmärkten-eine explorative Untersuchung am Beispiel des deutschen Marktes für persönliche Daten 2199-

- 8914, Forum Privatheit und selbstbestimmtes Leben in der digitalen Welt, Karlsruhe
- Buchanan JM. 1965. An Economic Theory of Clubs. *Economica* 32: 1-14
- BVDW. 2018. Datenwertschöpfung und Qualität von Daten, Bundesverband Digitale Wirtschaft (BVDW) e.V., Düsseldorf
- Caillaud B, Jullien B. 2003. Chicken & Egg: Competition Among Intermediation Service Providers. *The RAND Journal of Economics* 34: 309-28
- Calvino F, Criscuolo C, Marcolin L, Squicciarini M. 2018. A taxonomy of digital intensive sectors, Organisation for Economic Co-operation and Development, Paris
- Cao L. 2017. Data science: a comprehensive overview. *ACM Computing Surveys (CSUR)* 50: 43
- Capurro R. 1978. *Information. Ein Beitrag zur etymologischen und ideengeschichtlichen Begründung des Informationsbegriffs*. München: Saur Verlag.
- Caracciolo C, Aubin S, Whitehead B, Zervas P. *Research Conference on Metadata and Semantics Research2018*: 340-45. Springer.
- Carrière-Swallow Y, Haksar V. 2019. The Economics and Implications of Data - An Integrated Perspective, Washington, DC
- Casadesus-Masanell R, Hervas-Drane A. 2015. Competing with privacy. *Management Science* 61: 229-46
- Casadesus-Masanell R, Ruiz-Aliseda F. 2009. Platform competition, compatibility, and social efficiency. In *Working Paper*. Harvard Business School
- Checkland P. 1999. Systems thinking. *Rethinking management information systems: An interdisciplinary perspective*: 45-56
- Chellappa RK, Sin RG. 2005. Personalization versus privacy: An empirical examination of the online consumer's dilemma. *Information Technology and Management* 6: 181-202
- Chen M, Mao S, Liu Y. 2014. Big data: A survey. *Mobile networks and applications* 19: 171-209
- Cichy C, Rass S. 2019. An Overview of Data Quality Frameworks. *IEEE Access* 7: 24634-48
- Cornes R, Sandler T. 1986. *The Theory of Externalities, Public Goods, and Club Goods*. Cambridge, MI: Cambridge University Press.
- Crémer J, de Montjoye Y-A, Schweitzer H. 2019. Competition policy for the digital era - Final report, European Commission - DG COMP, Brussels
- Cuquet M, Fensel A. 2018. The societal impact of big data: A research roadmap for Europe. *Technology in Society* 54: 74-86
- Curry E. 2016. The Big Data Value Chain: Definition, Concepts, and Theoretical Approaches In *New Horizons for a Data-Driven Economy - A Roadmap for Usage and Exploitation of Big Data in Europe*, ed. JM Cavanillas, E Curry, W Wahlster, pp. 29-38: SpringerOpen
- Davenport TH, Barth P, Bean R. 2012. How 'big data' is different. *MIT Sloan Management Review* 54: 22-24
- Davis M, Martinez R, Kalaboukis C. 2010. Rethinking Personal Information – Workshop Pre-read, Invention Arts and World Economic Forum, Cologne
- de Mauro A, Greco M, Grimaldi M. 2016. A Formal Definition of Big Data based on its Essential Features. *Library Review* 65: 122-35
- Dersten S, Axelsson J, Fröberg J. 2011. *Effect Analysis of the Introduction of AUTOSAR: A Systematic Literature Review*. 239-46 pp.
- Dhar V. 2013. Data science and prediction. *Communications of the ACM* 56: 64-73
- Dosis A, Sand-Zantman W. 2019. The ownership of data. *Toulouse School of Economics. Working Paper*
- Drucker J. 2011. Humanities approaches to graphical display. *Digital Humanities Quarterly* 5: 1-21

- Duch-Brown N, Martens B, Mueller-Langer F. 2017. The economics of ownership, access and trade in digital data. JRC Digital Economy Working Paper 2017-01, Joint Research Centre, Seville
- Egan E. 2019. Charting a Way Forward: Data Portability and Privacy, Facebook
- European Commission. 2017. Communication from the Commission to the European Economic and Social Committee and the Committee of the Regions - "Building a European Datenökonomie" (SWD(2017) 2 final), European Commission, Brussels
- European Commission, IMF, OECD, UN, World Bank. 2009. System of National Accounts 2008, European Commission, International Monetary Fund, Organisation for Economic Co-operation and Development, United Nations and the World Bank, New York, NY
- Evans DS, Noel M. 2005. Defining Antitrust Markets When Firms Operate Two-Sided Platforms. *Columbia Business Law Review* 3: 101-34
- Evans DS, Schmalensee R. 2007. Antitrust Analysis of Multi-sided Platforms: The Industrial Organization of Markets with Two-sided Platforms. *Competition Policy International* 3: 151-79
- Fan J, Han F, Liu H. 2014. Challenges of Big Data analysis. *National Science Review* 1: 293-314
- Farboodi M, Veldkamp L. 2019. A Growth Model of the Datenökonomie, Working Paper, Columbia Business School, New York, June 20, New York, NY
- Fast V, Schnurr D, Wohlfarth M. 2019. *Data-driven Market Power: An Overview of Economic Benefits and Competitive Advantages from Big Data Use*. Presented at 47th Research Conference on Communications, Information, and Internet Policy (TPRC), Washington, DC
- Federico G, Morton FS, Shapiro C. 2019. Antitrust and Innovation: Welcoming and Protecting Disruption In *Innovation Policy and the Economy, Volume 20*: University of Chicago Press
- Feijóo C, Gómez-Barroso J-L, Aggarwal S. 2016. Economics of Big Data In *Handbook on the Economics of the Internet*, ed. JM Bauer, M Latzer, pp. 510-25. Cheltenham: Edward Elgar Publishing
- Feld H. 2019. The Case for the Digital Platform Act: Market Structure and Regulation of Digital Platforms, Roosevelt Institute
- Floridi L. 2010. Information: a very short guide. Oxford University Press, Oxford
- Frické M. 2009. The knowledge pyramid: a critique of the DIKW hierarchy. *Journal of information science* 35: 131-42
- Frické M. 2015. Big Data and its Epistemology. *Journal of the Association for Information Science and Technology* 66: 651-61
- FTC. 2014. Data Brokers: A Call for Transparency and Accountability, Washington, DC
- Furman J, Coyle D, Fletcher A, McAuley D, Marsden P. 2019. Unlocking digital competition - Report of Digital Competition Expert Panel, HM Treasury, London
- Furner J. 2016. "Data": The data In *Information Cultures in the Digital Age - A Festschrift in Honor of Rafael Capurro*, ed. M Kelly, J Bielby, pp. 287-306. Wiesbaden: Springer
- Gandomi A, Haider M. 2015. Beyond the hype: Big data concepts, methods, and analytics. *International journal of information management* 35: 137-44
- Gasser U, Palfrey J. 2007. Breaking Down Digital Barriers - When and How ICT Interoperability Drives Innovation
- Gitelman L. 2013. "Raw data" is an oxymoron. Cambridge, MA: MIT press.
- Goldfarb A, Tucker C. 2019. Digital Economics. *Journal of Economic Literature* 51: 3-43
- Goldfarb A, Tucker CE. 2011a. Online Display Advertising: Targeting and Obtrusiveness. *Marketing Science* 30: 389-404

- Goldfarb A, Tucker CE. 2011b. Privacy Regulation and Online Advertising. *Management Science* 57: 57-71
- Graef I, Wahyuningtyas SY, Valcke P. 2015. Assessing Data Access Issues in Online Platforms. *Telecommunications Policy* 39: 375-87
- Granados N, Gupta A. 2013. Transparency strategy: competing with information in a digital world. *MIS quarterly* 37: 637-41
- Grover V, Chiang RHL, Liang T-P, Zhang D. 2018. Creating Strategic Business Value from Big Data Analytics: A Research Framework. *Journal of Management Information Systems* 35: 388-423
- GSMA. 2018. The Data Value Chain, GSMA
- Günther WA, Mehrizi MHR, Huysman M, Feldberg F. 2017. Debating big data: A literature review on realizing value from big data. *The Journal of Strategic Information Systems* 26: 191-209
- Haav H-M, Küngas P. 2013. Semantic data interoperability: the key problem of big data. *Big Data Computing*: 245
- Hagiu A. 2007. Merchant or two-sided platform? *Review of Network Economics* 6: 115-33
- Haucap J. 2019. Competition and Competition Policy in a Data-Driven Economy. *Intereconomics*: 201-08
- Haucap J, Stühmeier T. 2016. Competition and Antitrust in Internet Markets In *Handbook on the Economics of the Internet*, ed. JM Bauer, M Latzer, pp. 183-210. Cheltenham, UK: Edward Elgar Publishing
- Haug A, Zachariassen F, Liempd D. 2011. The costs of poor data quality. *Journal of Industrial Engineering and Management* 4: 168-93
- Heinecke H, Schnelle K-P, Fennel H, Bortolazzi J, Lundh L, et al. 2004. Automotive open system architecture-an industry-wide initiative to manage the complexity of emerging automotive e/e-architectures, SAE Technical Paper
- Hestness J, Narang S, Ardalani N, Diamos G, Jun H, et al. 2017. Deep learning scaling is predictable, empirically. *arXiv preprint arXiv:1712.00409*
- Heumann S, Jentzsch N. 2019. Wettbewerb um Daten. Über Datenpools zu Innovationen. Berlin: Stiftung Neue Verantwortung e. V.
- Hey T, Tansley S, Tolle K. 2009. *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Redmont, WA: Microsoft Research.
- Hildebrandt C, Arnold R. 2016. Big Data und OTT-Geschäftsmodelle sowie daraus resultierende Wettbewerbsprobleme und Herausforderungen bei Datenschutz und Verbraucherschutz - WIK-Diskussionsbeitrag Nr. 414, Wissenschaftliches Institut für Infrastruktur und Kommunikationsdienste (WIK), Bad Honnef
- Hinz O, Eckert J. 2010. The impact of search and recommendation systems on sales in electronic commerce. *Business & Information Systems Engineering* 2: 67-77
- Hjørland B. 2019. Data (with Big Data and Database Semantics). *KO Knowledge Organization* 45: 685-708
- Ho SY, Bodoff D, Tam KY. 2011. Timing of adaptive web personalization and its effects on online consumer behavior. *Information Systems Research* 22: 660-79
- Hogan O, Holdgate L, Jayasuriya R. 2016. The Value of Big Data and the Internet of Things to the UK Economy, Cebr, London
- Horak R. 2008. *Telecommunications and data communications handbook*. Hoboken: John Wiley & Sons.
- IDC, Open Evidence. 2017. European Data Market SMART 2013/0063 - Final Report. A study prepared for the European Commission, IDC, Open Evidence
- Inmon WH. 2006. DW 2.0; Architecture for the Next Generation of Data Warehousing. *Information Management* 16: 8
- Inmon WH, Strauss D, Neushloss G. 2010. *DW 2.0: The architecture for the next generation of data warehousing*. Amsterdam: Elsevier.

- Jensen HE. 1950. Editorial Note In *Through Values to Social Interpretation: Essays on Social Contexts, Actions, Types and Prospects*, ed. H Becker, pp. vii-xi. Durham, NC: Duke University Press
- Jentzsch N, Sapi G, Suleymanova I. 2013. Targeted Pricing and Customer Data Sharing Among Rivals. *International Journal of Industrial Organization* 31: 131-44
- Jones CI, Tonetti C. 2019. Nonrivalry and the Economics of Data 0898-2937, National Bureau of Economic Research
- Junqué de Fortuny E, Martens D, Provost F. 2013. Predictive Modeling with Big Data: Is Bigger Really Better? *Big Data* 1: 215-26
- Kaase M. 2001. Databases, Core: Political Science and Political Behavior In *International Encyclopedia of the Social and Behavioral Sciences*, ed. NJ Smelser, PB Baltes, pp. 3251-55. Amsterdam: Elsevier
- Karwatzki S, Dytynko O, Trenz M, Veit D. 2017. Beyond the Personalization–Privacy Paradox: Privacy Valuation, Transparency Features, and Service Personalization. *Journal of Management Information Systems* 34: 369-400
- Katz ML, Shapiro C. 1985. Network Externalities, Competition, and Compatibility. *The American Economic Review* 75: 424-40
- Kempermann H, Lichtblau K. 2012. Definition und Messung von hybrider Wertschöpfung. *IW Trends* 39: 1-20
- Kim T, Barasz K, John LK. 2019. Why Am I Seeing This Ad? The Effect of Ad Transparency on Ad Effectiveness. *Journal of Consumer Research* 45: 906-32
- Kimball R. 2011. The evolving role of the enterprise data warehouse in the era of big data analytics - White Paper
- Kitchin R. 2014. *The data revolution: Big data, open data, data infrastructures and their consequences*. London: Sage.
- Krämer J, Wohlfarth M. 2018. Market power, regulatory convergence, and the role of data in digital markets. *Telecommunications Policy* 42: 154-71
- Kütt A, Priisalu J. *Proceedings of the International Conference on e-Learning, e-Business, Enterprise Information Systems, and e-Government (EEE)2014*: 1. The Steering Committee of The World Congress in Computer Science, Computer
- Lam WMW, Liu X. 2018. Does Data Portability Facilitate Entry?
- Lambrecht A, Tucker CE. 2013. When Does Retargeting Work? Information Specificity in Online Advertising. *Journal of Marketing Research* 50: 561-76
- Lambrecht A, Tucker CE. 2015. Can Big Data Protect a Firm from Competition? : SSRN
- Laporte S. 2018. Ideal language. *KO KNOWLEDGE ORGANIZATION* 45: 586-608
- Lee D. 2019. Big Data Quality Assurance Through Data Traceability: A Case Study of the National Standard Reference Data Program of Korea. *IEEE Access* 7: 36294-99
- Lee I. 2017. Big data: Dimensions, evolution, impacts, and challenges. *Business Horizons* 60: 293-303
- Lenart M, Bielecki A, Lesot M-J, Petrisor T, d'Allonnes AR. *International Conference on Artificial Intelligence and Soft Computing2018*: 579-91. Springer.
- Lerner AV. 2014. The Role of 'Big Data' in Online Platform Competition. SSRN 2482780
- Lewis RA, Rao JM. 2015. The unfavorable economics of measuring the returns to advertising. *The Quarterly Journal of Economics* 130: 1941-73
- Li X, Ling CX, Wang H. 2016. The convergence behavior of naive Bayes on large sparse datasets. *ACM Transactions on Knowledge Discovery from Data (TKDD)* 11: 10:1-10:24
- Lichtblau K, Arnold R. 2012. Smart Industry – Intelligente Industrie: Eine neue Betrachtungsweise der Industrie. Ergebnisse einer Studie der Institut der

- deutschen Wirtschaft Köln Consult GmbH für das Land Hessen, Initiative Industrieplatz Hessen, Neu-Isenburg
- Lind H-G, Suckfüll H. 2013. Die Initiative zu einer Deutschen Daten Treuhand (DEDATE) als ultima ratio der persönlichen digitalen Datenwirtschaft (PDD) - Ansätze und Strukturen für eine gezielte Verwertung persönlicher Daten unter Berücksichtigung aller Interessengruppen - Dateneigentümer, Wirtschaft und Staat, Fraunhofer, HSBD GmbH, Leipzig
- Liu J, Li J, Li W, Wu J. 2016. Rethinking big data: A review on the data quality and usage issues. *ISPRS journal of photogrammetry and remote sensing* 115: 134-42
- Liu Q, Serfes K. 2006. Customer information sharing among rival firms. *European Economic Review* 50: 1571-600
- Locks O, Winkler G. 2017. *Future Vehicle System Architecture*. Presented at ASAM General Assembly,
- Lundvall B-Å, Borrás S. 2005. Science, technology and innovation policy In *The Oxford handbook of innovation*, ed. J Fagerberg, DC Mowery, RR Nelson, pp. 599-631. Oxford: Oxford University Press
- Mahnke RP. 2015. Big Data as a Barrier to Entry. *Antitrust Chronicle* 12: 1-6
- Manyika J, Chui M, Groves P, Steve F, Kuiken V, Doshi EA. 2013. Open data: Unlocking innovation and performance with liquid information, McKinsey Global Institute
- Marshall P. 2012. What you need to know about big data. *Government Computer News*, 07. February 2012:
- Martens B. 2016. An Economic Policy Perspective on Online Platforms, European Commission
- McKinsey. 2016. The age of analytics: Competing in a data-driven world
- Mirzaie M, Behkamal B, Paydar S. 2019. Big Data Quality: A systematic literature review and future research directions. *arXiv preprint arXiv:1904.05353*
- Moore FT. 1959. Economies of scale: Some statistical evidence. *The Quarterly Journal of Economics* 73: 232-45
- Morton FS, Bouvier P, Ezrachi A, Jullien B, Katz R, et al. 2019. Stigler Committee on Digital Platforms - Final report, Stigler Center
- Muschalle A, Stahl F, Löser A, Vossen G. *international workshop on business intelligence for the real-time enterprise2012*: 129-44. Springer.
- Nissenbaum H. 2011. A Contextual Approach to Privacy Online. *Daedalus* 140: 32-48
- Norberg PA, Horne DR, Horne DA. 2007. The Privacy Paradox: Personal Information Disclosure Intentions versus Behaviors. *Journal of Consumer Affairs* 41: 100-26
- O'Mahony MP, Hurley NJ, Silvestre G. *Proceedings of the 11th international conference on Intelligent user interfaces2006*: 109-15. ACM.
- OECD. 1996. The Essential Facilities Concept, Paris
- OECD. 2015. The OECD Model Survey on ICT Usage by Businesses - 2nd revision, Organization for Economic Co-operation and Development, Paris
- OECD. 2016. Big Data: Bringing Competition Policy to the Digital Era, OECD
- OECD. 2018. Rethinking Antitrust Tools for Multi-Sided Platforms, OECD
- OECD, Eurostat. 2018. *Oslo Manual 2018: Guidelines for Collecting, Reporting and Using Data on Innovation, 4th Edition, The Measurement of Scientific, Technological and Innovation Activities*. Paris and Luxembourg: OECD Publishing and Eurostat.
- Ohm P. 2010. Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization. *UCLA Law Review* 57: 1701-77
- Otto B, Auer S, Cirullies J, Jürjens J, Menz N, et al. 2016. Industrial data space: digital sovereignty over data, Fraunhofer, München

- Otto B, Lohmann S, Steinbuß S, Teuscher A. 2018. IDS reference architecture model, industrial data space, version 2.0. *International Data Spaces Association & Fraunhofer*
- Panzar JC, Willig RD. 1981. Economies of scope. *The American Economic Review* 71: 268-72
- Parker GG, van Alstyne MW. 2005. Two-sided network effects: A theory of information product design. *Management science* 51: 1494-504
- Pauer A, Nagel L, Fedkenhauser T, Fritzsche-Sterr Y, Resetko A. 2018. Data exchange as a first step towards Datenökonomie, London
- Paunov C, Planes-Satorra S. 2019. How are digital technologies changing innovation? Evidence from agriculture, the automotive industry and retail, Organisation for Economic Co-operation and Development, Paris
- Peppers D, Rogers M, Dorf B. 1999. Is your company ready for one-to-one marketing. *Harvard Business Review* 77: 151-60
- Pine BJ, Peppers D, Rogers M. 1995. *Do you want to keep your customers forever?* Boston: Harvard Business Press.
- Pipino LL, Lee YW, Wang RY. 2002. Data quality assessment. *Communications of the ACM* 45: 211-18
- Planes-Satorra S, Paunov C. 2019. The Digital Innovation Policy Landscape in 2019. OECD Science, Technology and Innovation Policy Papers No. 71, Organisation for Economic Co-operation and Development, Paris
- Porter ME. 1985. *Competitive Advantage: Creating and sustaining superior performance*. New York: The Free Press.
- Prüfer J, Schottmüller C. 2017. Competing with big data. *Tilburg Law School Research Paper*
- Qi Z, Wang H, Li J, Gao H. 2018. Impacts of dirty data: and experimental evaluation. *arXiv preprint arXiv:1803.06071*
- Redman TC. 1998. The impact of poor data quality on the typical enterprise. *Communications of the ACM* 41: 79-82
- Redman TC. 2016. Bad data costs the US \$3 trillion per year. *Harvard Business Review*, 22. September 2016:
- Rocher L, Hendrickx JM, de Montjoye Y-A. 2019. Estimating the success of re-identifications in incomplete datasets using generative models. *Nature Communications* 10: 3069
- Rochet JC, Tirole J. 2003. Platform Competition in Two-Sided Markets. *Journal of the European Economic Association* 1: 990-1029
- Roman D, Stefano G. 2016. *Towards a Reference Architecture for Trusted Data Marketplaces: The Credit Scoring Perspective*. Presented at 2nd International Conference on Open and Big Data (OBD), Vienna
- Rubinfeld DL, Gal MS. 2017. Access Barriers to Big Data. *Arizona Law Review* 59
- Rysman M. 2009. The Economics of Two-Sided Markets. *The Journal of Economic Perspectives* 23: 125-43
- Salgado D, Esteban E, Saldana S, Oancea B, Sakarovitch B, et al. 2018. *Estimation of population counts combining official data and aggregated mobile phone data*. Presented at European Conference on Quality in Official Statistics, Kraków
- Salinas SO, Lemus AC. 2017. Data warehouse and big data integration. *Int. Journal of Comp. Sci. and Inf. Tech* 9: 1-17
- Samat S, Acquisti A, Babcock L. *Thirteenth Symposium on Usable Privacy and Security (SOUPS 2017), Santa Clara, 2017: 299-319*.
- Saudagar M, Ye M, Al-Otaibi S, Al-Jarba K. 2019. Smart Manufacturing: Hope or Hype? *CEP* 115: 43-48
- Schafer JB, Konstan JA, Riedl J. 2001. E-commerce recommendation applications. *Data Mining and Knowledge Discovery* 5: 115-53

- Schepp N-P, Wambach A. 2016. On Big Data and its Relevance for Market Power Assessment. *Journal of European Competition Law & Practice* 7: 120-24
- Schroeder R. 2016. Big data business models: Challenges and opportunities. *Cogent Social Sciences* 2
- Schumann JH, von Wangenheim F, Groene N. 2014. Targeted online advertising: Using reciprocity appeals to increase acceptance among users of free web services. *Journal of Marketing* 78: 59-75
- Schwartz R, Dodge J, Smith NA, Etzioni O. 2019. Green AI. *arXiv preprint arXiv:1907.10597*
- Schweitzer H. 2019. Datenzugang in der Datenökonomie: Eckpfeiler einer neuen Informationsordnung. *GRUR* 121: 569-80
- Shapiro C, Varian HR. 1999. *Information Rules. A Strategic Guide to the Network Economy*. Boston, MA: Harvard Business School Press.
- Sidak JG, Lipsky AB. 1999. Essential Facilities. *Stanford Law Review* 51: 1187-249
- Sidi F, Panahy PHS, Affendey LS, Jabar MA, Ibrahim H, Mustapha A. 2012. *International Conference on Information Retrieval & Knowledge Management 2012*: 300-04. IEEE.
- Sinha R, Swearingen K. *CHI'02 Extended Abstracts on Human Factors in Computing Systems, Minneapolis, 2002*: 830-31.
- Srai JS, Settanni E, Tsolakis N, Aulakh PK. 2019. *Supply Chain Digital Twins: Opportunities and Challenges Beyond the Hype*. Presented at 23rd Cambridge International Manufacturing Symposium, Cambridge
- Stuckenschmidt H. 2012. *Data semantics on the web*. Springer
- Swire P, Lagos Y. 2013. Why the Right to Data Portability Likely Reduces Consumer Welfare: Antitrust and Privacy Critique. *Maryland Law Review* 72: 353-80
- Taleb I, Dssouli R, Serhani MA. *2015 IEEE international congress on big data 2015*: 191-98. IEEE.
- Taleb I, Serhani MA, Dssouli R. *2018 International Conference on Innovations in Information Technology (IIT) 2018*: 69-74. IEEE.
- Tam KY, Ho SY. 2006. Understanding the impact of web personalization on user information processing and decision outcomes. *MIS Quarterly* 30: 865-90
- Tao F, Zhang H, Liu A, Nee AY. 2018. Digital twin in industry: state-of-the-art. *IEEE Transactions on Industrial Informatics* 15: 2405-15
- Thirumalai S, Sinha KK. 2013. To personalize or not to personalize online purchase interactions: implications of self-selection by retailers. *Information Systems Research* 24: 683-708
- Tiefenbacher K, Olbrich S. *ECIS, Münster, Germany, 2015*.
- Tsai JY, Egelman S, Cranor L, Acquisti A. 2011. The effect of online privacy information on purchasing behavior: An experimental study. *Information Systems Research* 22: 254-68
- Tucker CE. 2010. The Economic Value of Online Customer Data. *Economics of Personal Data and Privacy* 30
- Tucker CE. 2014. Social networks, personalized advertising, and privacy controls. *Journal of Marketing Research* 51: 546-62
- Tucker DS, Wellford HB. 2014. Big Mistakes Regarding Big Data. In *Antitrust Source, American Bar Association*
- Wang RY, Strong DM. 1996. Beyond accuracy: What data quality means to data consumers. *Journal of management information systems* 12: 5-33
- Wang W, Xu J, Wang M. 2018. Effects of recommendation neutrality and sponsorship disclosure on trust vs. distrust in online recommendation agents: Moderating role of explanations for organic recommendations. *Management Science* 64: 5198-219

- Wiengarten L, Zwick M. 2018. Neue digitale Daten in der amtlichen Statistik. *WISTA* 2017: 43-60
- Wohlfarth M. 2019. Data Portability on the Internet. *Business & Information Systems Engineering* 61: 551-74
- World Economic Forum. 2011. Personal Data: The Emergence of a New Asset Class, WEF, Cologny
- Zins C. 2007. Conceptual approaches for defining data, information, and knowledge. *Journal of the American society for information science and technology* 58: 479-93