2.1 Kommerzielle Datenmärkte

Abstract: Der kommerzielle Handel mit Daten hat sich im Rahmen von Digitalisierung und digitaler Transformation als immer wichtigerer Wirtschaftsbereich etabliert, in dem Datenmarktplätzen als Handelsplattformen eine Schlüsselrolle zukommt. Es werden mittlerweile neue Geschäftsmodelle identifiziert, die von unterschiedlichen Playern entwickelt und betrieben werden und die ausschließlich auf einer Aggregation und anschließenden Nutzung von Daten basieren. Dieses Kapitel gibt einen Überblick über die Funktionalität von Datenmarktplätzen mit und ohne Feedback-Loop und zeigt deren Entwicklung auf. Dabei betrachten wir einerseits den „klassischen“ Ansatz, bei dem große Datensammlungen angelegt, eventuell aufbereitet und sodann ganz oder teilweise zum Kauf angeboten werden; in diesem Bereich kommen zahlreiche Techniken der Daten-Vorverarbeitung oder des Data Cleansing zur Anwendung; ferner hat sich aktuelle Forschung z. B. mit der Frage der Preisbildung beschäftigt. Später betrachten wir den an die Plattformökonomie angelehnten Ansatz, bei dem sich die gesammelten oder anfallenden Daten durch Einsatz von Feedback-Loops selbst aufwerten. Auch wird die Frage betrachtet, inwieweit die einzelne Person vom Handel mit ihren Daten Kenntnis erlangen kann, wie sie ihn unterbinden bzw. wie sie selbst Nutzen daraus ziehen kann.

Einleitung

Informationen zählen heute zu einer der wichtigsten Ressourcen unserer Gesellschaft, da sie u. a. die Verbreitung und Erzeugung von neuem Wissen ermöglichen. In Form von strukturierten, aufbereiteten Daten bieten Informationen einen Wert für ein breites Spektrum an Konsumierenden, die durch die Analyse von Daten Unterstützung bei Geschäftsentscheidungen erhalten können. Außerdem können Daten als Grundlage für Dienstleistungen verwendet oder nach einer angemessenen Verarbeitung weiterverkauft werden. Im Zeitalter der umfassenden und inzwischen sämtliche Bereiche des täglichen Lebens erreichenden Digitalisierung und digitalen Transformation entstehen daher seit mehreren Jahren Marktplätze für Daten in gewisser Analogie zu Marktplätzen (Börsen) für Aktien, Strom oder andere Waren. Wesentlich ist dabei nicht nur der eigentliche Handel mit Daten, sondern deren (Vor-) Verarbeitung, Konsolidierung, Anreicherung und insbesondere Analyse, etwa unter Verwendung von Information-Retrieval-Techniken, Natural-Language-Processing (NLP) oder statistischen Methoden. Anwendungen finden sich u. a. im Marketing, in der Finanzindustrie, im Gesundheitswesen oder in der Unternehmenssteuerung (Business Intelligence). Insbesondere größere Unternehmen und Fachabteilungen

in Konzernen verfügen oft über spezifisches Wissen für eine Übersetzung von Daten in monetär verwertbare Informationen, bspw. Fachwissen in Bezug auf eine spezielle Kunden- oder Produktnische oder Beziehungen zu potenziellen Kundinnen oder Kunden, die bereit sind, für diese Informationen zu bezahlen.

Zur Analyse von Daten sind im Allgemeinen IT-Infrastrukturen notwendig, die Daten aus dem Web oder anderen Quellen sammeln und mit Informationen über ihre Herkunft, ihren Inhalt und ihre sprachliche Struktur anreichern. In der Regel erfordern derartige Unterfangen eine erhebliche Infrastruktur mit entsprechenden Investitionen und erzeugen weitere Kosten für die Aktualisierung von Daten, welche sich insbesondere Mittelständler und Fachabteilungen oft nicht leisten können. Abhilfe schaffen hier die Einrichtung und der Betrieb einer Plattform, die bei hinreichendem Kundeninteresse in der Lage ist, durch einen Feedback-Loop ihre Daten selbst weiter anzureichern und damit deren Wert zu verbessern.

Dieses Kapitel soll eine Übersicht geben über den Stand bei kommerziellen Datenmarktplätzen. Es zeigt deren Aufbau auf und geht auf unterschiedliche Arten von Datenmarktplätzen ein. Dazu behandeln wir in Abschnitt 1 zunächst Datenmarktplätze allgemein mit den Schwerpunkten technischer Aufbau und Preisbildung und stellen unterschiedliche Klassen von Datenmarktplätzen vor. Abschnitt 2 behandelt Datenmarktplätze in der Plattform-Ökonomie, durch welche Feedback-Loops sowie Techniken der künstlichen Intelligenz einbezogen und genutzt werden. Abschnitt 3 beschließt unsere Ausführungen.

1 Datenmarktplätze

Mit Daten wurde spätestens seit der Erfindung des Telefons gehandelt, etwa über Auskunftsteien, aber wahrscheinlich ist der Handel mit Daten sogar älter. Durch den Übergang von analogen zu digitalen Daten im Rahmen der weltweit betriebenen Digitalisierung wurde dies grundsätzlich vereinfacht, wenn auch mit neuen Herausforderungen versehen, und hat in der Folge rapide zugenommen. Bereits in den Anfangsjahren des Internets gab es sog. „Information Broker“, deren Job es war, zu bestimmten Themen, die von Auftraggebern vorgegeben wurden, Informationen im Web zu sammeln, zu konsolidieren und in angemessener Form zu präsentieren. De facto sind Informationsvermittler (laut Wikipedia) „privatwirtschaftliche Unternehmen, die gegen ein Honorar die Recherche von Informationen übernehmen“;¹ es gibt sogar laut Bundesagentur für Arbeit eine Ausbildung zur Informations-Brokerin bzw. zum Informations-Broker. Unsere Darstellung in diesem Abschnitt folgt der

1 <https://de.wikipedia.org/wiki/Informationsvermittler>. Letztes Abrufdatum der Internet-Dokumente ist der 15.11.2020.

von Lange et al.² Durch die fortschreitende Digitalisierung und Automatisierung von Anwendungen sind aus Brokern an vielen Stellen Marktplätze geworden; von diesen ist im Folgenden die Rede.

1.1 Aufbau von Datenmarktplätzen

Datenmarktplätze bieten eine Plattform für den Handel mit Daten als Informationsgut im elektronischen Markt und sind grundsätzlich aufgebaut wie in Abb. 1 dargestellt. Der eigentliche Marktplatz umfasst neben Daten auch Algorithmen zu deren Bereinigung, Veredelung, Aggregation usw., d. h. zur Anwendung algorithmischer Prozeduren, die von Entwicklerinnen oder Entwicklern bereitgestellt werden. Neben denjenigen, die Algorithmen entwickeln, greifen auch diejenigen, die Anwendungen entwickeln, auf den Marktplatz zu, um Daten für die von ihnen entwickelten Anwendungen zu erhalten. Neben Anwendungsentwicklern sind die in Abb. 1 gezeigten Analysten weitere Kunden, die Daten gegen Geld beziehen und zumeist in ihre eigenen Anwendungen integrieren. Auf der anderen Seite fließt Geld an die Datenlieferanten sowie an die Algorithmen-Entwickler.

Es sei bemerkt, dass Datenmarktplätze zahlreiche Preismodelle aufweisen (s. Abschnitt 1.3). Grundsätzlich ist die Kostenstruktur von Daten eine besondere Herausforderung, da die Fixkosten in Form von verhältnismäßig hohen, anfänglichen Produktions- und Sammelkosten gegenüber den variablen Kosten etwa einer Vervielfältigung stark dominieren. Preismodelle für Daten von Datenmarktplätzen umfassen daher Ansätze wie Free, Free + Premium = Freemium, Pay-per-Use, Flatrate oder Mitgliedschaft sowie Kombinationen dieser.

Abb. 1: Grundsätzlicher Aufbau eines Datenmarktplatzes

² Vgl. Lange et al. 2018.

Abb. 2 zeigt ein detaillierteres Bild eines Datenmarktplatzes, in dem die einzelnen Akteurinnen und Akteure genauer spezifiziert werden.³ Der Kern des Marktplatzes, nach Abb. 1 bestehend aus Daten und Algorithmen, untergliedert sich meist in mehrere Ebenen, die einerseits die Verarbeitungsinfrastruktur umfassen und andererseits die entweder bereitgestellte oder zugeliferte Funktionalität.

Daten können aus öffentlichen oder nicht-öffentlichen Quellen stammen; Interaktionen (sowie Geldflüsse) finden mit unterschiedlichen Benutzergruppen statt.⁴

Abb. 2: Detailliertere Sicht auf einen Datenmarktplatz

Die Hauptakteure eines Datenmarktplatzes lassen sich in die drei Gruppen unterteilen: Marktplatzbetreibende, diejenigen, die Daten anbieten bzw. verkaufen, sowie diejenigen, die Daten kaufen bzw. nutzen. Datenmarktplatzbetreibende fungieren als Intermediär zwischen den Kaufenden und Verkaufenden. Ihre Hauptaufgabe besteht dabei im Sammeln der Daten der Anbietenden und dem Verkauf von Daten über Datenabfragen. Datenanbietende besitzen Daten, die sie den Datenmarktbetreibenden umsonst, gegen Bezahlung oder gegen eine andere Form der Entschädigung zur Verfügung stellen; hierbei kann es sich um private Personen handeln, jedoch wird zumeist unterschieden zwischen Anbietenden gewerblicher und nicht gewerblicher Daten. Daten kauft, wer, wie erwähnt, ein Daten analysiert oder Anwendungen entwickelt. Ein Funktionsreferenzmodell für Datenmarktplätze, welche

³ Vgl. Muschalle et al. 2012, 132.

⁴ Weitere Einzelheiten beschreiben Muschalle et al. 2012.

zentrale Funktionen eines Datenmarktplatzes anhand der Funktionsgruppen Schnittstellen und Sicherheit, Data Service Ecosystem, Datenintegration und Administration ordnet, wurde von Meisel und Spiekermann entworfen.⁵

1.2 Klassen von Datenmarktplätzen

Die konzeptionelle Betrachtung von Datenmarktplätzen im Hinblick auf deren Charakteristika findet sich insbesondere in einer Reihe von Studien, die von Schomm, Stahl und Vossen durchgeführt wurden.⁶ Hierbei werden Datenanbieterende und Datenmarktplätze in objektiven und subjektiven Dimensionen analysiert, was eine Charakterisierung und vergleichende Übersicht durch gezielte Analyse der einzelnen Kriterien ermöglicht. So lassen sich Datenmarktplätze in vier Hauptklassen

- kommerzielle Datenmarktplätze,
- Datenmarktplätze für persönliche Daten,
- Datenmarktplätze für öffentliche Daten und
- Schwarzmärkte für gestohlene Daten⁷

einteilen, was in Abb. 3 gezeigt ist.

Abb. 3: Klassen von Datenmarktplätzen (nach Lange et al. 2018)

⁵ Vgl. Meisel und Spiekermann 2019.

⁶ Vgl. Schomm, Stahl und Vossen 2013; Stahl, Schomm und Vossen 2014; Stahl et al. 2016.

⁷ S. z. B. <https://www.symantec.com/connect/blogs/underground-black-market-thriving-trade-stolen-data-malware-and-attack-services>.

Die erste und bisher häufigste Klasse der kommerziellen Datenmarktplätze beinhaltet jegliche Datenmarktplätze, die Daten von kommerziellen Datenanbietenden für kommerzielle sowie private Datennutzende bereitstellen. Auf beiden Handelsseiten stehen bei diesen Datenmarktplätzen häufig gewinnorientierte Unternehmen. Innerhalb dieser Klasse lassen sich zwei Unterkategorien bilden: kommerzielle Datenmarktplätze mit einem breit gefächerten und eher generellen Datenangebot und solche, die sich vollständig auf den Handel mit einer bestimmten Datenart spezialisiert haben. Unter die generellen kommerziellen Datenmarktplätze fallen Plattformen wie Microsoft Azure Marketplace, Statista oder auch Advaneo, da diese ein vielfältiges Angebot von unterschiedlichen Daten bieten. Die spezialisierten Datenmarktplätze können weiterhin nach ihrem Datenangebot in nahezu beliebig viele Themenfelder unterteilt werden. Ein Beispiel ist CoreLogic, ein Datenmarkt für Liegenschaften und Immobilien; ein weiteres ist Acxiom mit Daten für unterschiedlichste Marketing-Anwendungen. In diese Kategorie fällt auch Credit Sesame in China.

Die zweite Kategorie umfasst den Handel mit privaten und persönlichen Daten. Diese Daten werden von Privatpersonen angeboten; persönliche Daten sind hier also nicht die Kundendaten eines Unternehmens, sondern beziehen sich auf Informationen, die direkt von Privatpersonen preisgegeben (und idealerweise monetarisiert) werden. Diese bieten ihre Daten auf dem Datenmarktplatz gegen eine Entschädigung (zumeist in Form einer bestimmten Funktionalität oder einer Dienstleistung) an. Auf der Käuferseite stehen hier wie bei den kommerziellen Datenmarktplätzen hauptsächlich Unternehmen, deren Datenanalyse diese Daten für ihre Unternehmenszwecke, wie beispielsweise Marketingstrategien, gewinnbringend einsetzen möchten. Neben DataFairPlay und BitsAboutMe ist hier auch bridgit.io zu nennen.

Die Kategorie Datenmarktplätze für öffentliche Daten⁸ schließt alle Datenangebote ein, die kostenlos für jeden frei zugänglich sind. Dies beinhaltet vor allem die Datenbereitstellung durch Regierungen und öffentliche Verwaltungen,⁹ die ihre Daten wie beispielsweise statistische Auswertungen oder Protokolle im Internet zur Verfügung stellen. Auf der Nutzerseite dieser Datenmarktplätze finden sich neben rein privaten Nutzenden und Unternehmen auch kommerzielle Datenmarktplätze. Diese benutzen die kostenlosen Daten, um ihr eigenes Datenangebot zu vergrößern oder durch Kombination mit eigenen Daten den Datensätzen neuen Wert beizufügen, ein Vorgehen, das insbesondere von Start-ups häufig praktiziert wird. In beiden Fällen verbessert sich durch ein breites Datenangebot die Möglichkeit, neue Kundinnen und Kunden zu gewinnen. Die letzte Kategorie der Schwarzmärkte für Daten darf nicht unbeachtet bleiben, sofern der Markt und der Handel von Daten vollständig betrachtet werden sollen. Wie für viele andere Handelsgüter gibt es

⁸ S. a. Beitrag von Schieferdecker, Kap. 2.3 in diesem Praxishandbuch.

⁹ Wie das amerikanische Portal <https://www.data.gov/>, das britische <https://data.london.gov.uk/> oder das deutsche <https://www.govdata.de>.

auch für Daten illegale Handelsplattformen, auf denen unrechtmäßig beschaffte oder gesetzeswidrig gesammelte Daten verkauft werden.

Neben diesen Klassen von Datenmarktplätzen können auch unterschiedliche Arten des Datenhandels unterschieden werden: Letzterer kann verdeckt erfolgen, z. B., wenn mit Kundendaten, Logdaten oder Ähnlichem gehandelt wird, ohne dass diejenigen, die die Daten eigentlich erzeugen, davon Kenntnis erhalten. Verdeckter Handel liegt z. B. vor, wenn Google oder Facebook Nutzerdaten zum Advertising nutzen oder verkaufen. Demgegenüber liegt offener Datenhandel vor, wenn mit Daten über einen Marktplatz gehandelt wird, zu dem jeder Mensch Zugang hat, der die Eingangsbedingungen (sofern es solche gibt) erfüllt.

1.3 Preisbildung

Das Festsetzen von Preisen für Daten ist von zentraler Bedeutung für jeden Datenmarktplatz und zudem eine besondere Herausforderung. Wie bei anderen Gütern auch ist die Bestimmung des Wertes von Daten für die Käuferin bzw. den Käufer von essentieller Bedeutung für eine Preisfindung. Die Schwierigkeit liegt darin, dass dieser Wert zumeist nicht bekannt ist. Ebenfalls wie bei anderen Gütern hängt der Wert von Daten von der Marktsituation und dem Verhältnis von Angebot und Nachfrage ab. Demnach führt ein großes Datenangebot bei gleichzeitig geringer Nachfrage zu einem geringeren Wert der Daten und umgekehrt. Hinzu kommt, dass der Gebrauchswert von Daten, die ja auf einfache Weise beliebig vervielfältigbar sind, mit der Zeit abnimmt und dass der Wert von Informationsgütern subjektiv ist und von verschiedenen Personen als unterschiedlich hoch empfunden wird. Daher fehlt Datenanbietenden oftmals das Verständnis für den Wert ihrer Daten für Konsumierende, zumal eine zuverlässige Methode für die Bestimmung des Wertes von Datensätzen bisher nicht existiert. Durch den engen Zusammenhang des Wertes von Daten mit der Einschätzung der Konsumierenden ergibt sich, dass der Wert und damit auch der Preis von Daten in Abhängigkeit von den Nachfragenden bestimmt werden muss; Beispiele für entsprechende Ansätze liefern Stahl et al.¹⁰ sowie Stahl und Vossen.¹¹

Es hat verschiedene Versuche gegeben, den Wert insbesondere persönlicher Daten zu ermitteln. Ein bekanntes Beispiel ist ein von der Financial Times entwickelter Rechner,¹² mit der Privatpersonen vielleicht nicht den absoluten Wert ihrer persönlichen Daten ermitteln können, der aber zumindest zeigt, wie der Wert in Abhängigkeit von z. B. Familienstand oder Besitz allmählich steigt. Ein weiterer Ansatz zur

¹⁰ Vgl. Stahl et al. 2015.

¹¹ Vgl. Stahl und Vossen 2016; Stahl und Vossen 2017.

¹² S. <https://ig.ft.com/how-much-is-your-personal-data-worth/>.

groben Ermittlung des Werts persönlicher Daten ist der Vergleich der Preise, die bei Akquisitionen gezahlt werden (etwa die von WhatsApp durch Facebook oder die von LinkedIn durch Microsoft). Man vergleiche hierzu auch die Ergebnisse der Studie von Goldhammer und Wiegand 2017.

Preisfindung steht auch im Fokus verschiedener Forschungsaktivitäten. Sämtliche Anforderungen an Preismodelle lassen sich in die drei Kategorien logische, technische und ökonomische Anforderungen gliedern. Zu den logischen Anforderungen gehören all jene, deren Ziel ein logisch einwandfreies Preismodell darstellt. Hierzu zählt beispielsweise die Anforderung der Arbitrage-Freiheit, was bedeutet, dass für die Datenkäuferin bzw. den Datenkäufer keine Möglichkeit bestehen darf, die Antwort auf eine teure Datenanfrage durch ein Zusammensetzen von Antworten mehrerer günstigerer Anfragen zu erhalten. Die Forschung in diesem Bereich geht insbesondere zurück auf Koutris et al.¹³ Tang et al. gehörten zu den ersten, die Datenqualität in die Preisbildung einbezogen haben;¹⁴ diese Idee wurde von Stahl und Vossen aufgegriffen und weitergeführt.¹⁵ Insbesondere kann eine Benutzerin bzw. ein Benutzer dort zwischen unterschiedlichen berechenbaren Qualitätskriterien (wie Aktualität oder Vollständigkeit) wählen, diese gewichten und sie bzw. er erhält dann ein Datenprodukt, das diesen Wünschen entspricht. Unter die technischen Anforderungen an Preismodelle fallen jegliche Anforderungen, die die technische Umsetzung und Ausführung der Preismodelle adressieren, beispielsweise eine möglichst kurze Rechenzeit eines automatischen Preisfindungssystems oder einer Anfragebeantwortung. In der Literatur finden sich ferner Arbeiten zur Preisfindung, welche die ökonomischen Aspekte von Datenmarktplätzen, etwa die Fairness von Preismodellen einschließen.

Weiterhin lassen sich Preismodelle nach den Basiseinheiten der Daten unterscheiden; gängig sind einzelne Records, Datenobjekte oder Sichten (Views), wobei Tupel eine feinere Granularität als Views aufweisen. Ferner lassen sich Preismodelle in anfragebasierte und qualitätsbasierte Preismodelle untergliedern. Bei ersteren werden die Preise für Daten anhand der getätigten Datenbankanfragen durch Konsumierende bestimmt. Dabei existieren verschiedene Methoden, den Preis für eine Antwort auf eine Anfrage zu berechnen; häufig berechnen Preisfunktionen den Preis von Daten basierend auf festen Beträgen je konsumierter Basisdateneinheit, also je Datentupel oder je View. Hierfür muss das Preismodell bestimmen können, aus welchen Basiseinheiten die Antwort auf eine Anfrage zusammengesetzt wurde.¹⁶

¹³ Vgl. Koutris et al. 2015.

¹⁴ Vgl. Tang et al. 2014.

¹⁵ Vgl. Stahl und Vossen 2016; Stahl und Vossen 2017.

¹⁶ Für weitere Einzelheiten zu den einzelnen Preisfindungsansätzen vgl. Lange et al. 2018 und Stahl et al. 2015.

Interessanterweise wird in allen genannten formalen Preismodellen bisher lediglich für Daten, also Anfrageergebnisse ein Preis berechnet, nicht jedoch für die Ausführung der Berechnung selbst. In der Realität ist dies oft anders, denn wie in den Abb. 1 und 2 gezeigt umfasst ein Datenmarktplatz typischerweise auch die Berechnungsfunktionalität, deren Nutzung sich die Betreibenden bezahlen lassen. Ein erster Ansatz zur gleichzeitigen Betrachtung von Daten- und Berechnungskosten findet sich bei Martins et al.,¹⁷ einer Benutzerin bzw. einem Benutzer steht hier eine Reihe vorgegebener Anfragen zur Verfügung, die jeweils mit einem Preis versehen sind. Sie bzw. er kann diese Anfragen gewichten, um durch ihre Ausführung ein möglichst optimales (Daten-)Ergebnis zu erzielen als auch ein vorgegebenes Budget nicht zu überschreiten. Das Problem lässt sich als Knapsack-Optimierungsproblem¹⁸ formulieren, was zur Folge hat, dass auf approximative Algorithmen zu seiner Lösung zurückgegriffen werden muss.

2 Datenmarktplätze in der Plattform-Ökonomie

In diesem Abschnitt wird auf die neueren Entwicklungen im Bereich Datenmarktplätze eingegangen, die sich von den bisher Beschriebenen wesentlich durch das Vorhandensein eines Feedback-Loops unterscheiden. Dieser liefert dem Betreibenden mehr Daten, was bessere Analysen erlaubt, aus denen sich bessere Datenprodukte ergeben. Hieraus resultieren schließlich zufriedenere Kundinnen und Kunden, was sich wiederum in mehr Empfehlungen sowie in Neukundinnen und -kunden niederschlägt.

3.1 Datenmarktplätze und der Feedback-Loop

Idealerweise sollte ein Datenmarktplatz mit jeder Transaktion, z. B. einer verkauften Daten- oder Serviceeinheit einen besseren Service anbieten können und dadurch noch mehr Kundschaft anlocken können. Allerdings wissen diejenigen, die den Datenmarktplatz betreiben, oft nicht, was genau mit den Daten nach Kauf geschieht, also z. B. in welches maschinell gelernte Modell diese Daten einfließen. Fast immer haben zudem die Datenmarktplatzbetreibenden keine Rechte zur Verwertung des Modells.

¹⁷ Vgl. Martins et al. 2019.

¹⁸ Beim Knapsack- oder Rucksack-Problem geht es darum, für einen Rucksack mit begrenzter Kapazität Dinge zum Mitnehmen so auszuwählen, dass ihr Nutzen optimiert, aber die Kapazitätsbeschränkung eingehalten wird, vgl. https://en.wikipedia.org/wiki/Knapsack_problem.

Den Datenmarktplatzbetreibenden fehlt damit ein wichtiges Element, der „Feedback-Loop“, der den Betreibenden z. B. wertvolle Informationen über die Qualität bzw. den „Wert“ der angebotenen Daten für eine spezifische Branche geben könnte. Vielmehr implementieren aktuelle Datenmarktplätze oft noch immer das „Pipeline“-Modell eines Extract-Transform-Load-Prozesses (ETL) eines Data Warehouse für möglichst zahlreiche Kundinnen und Kunden. Sie skalieren so mit der Fähigkeit, neue Kundschaft zu gewinnen, die sich von der Verwendung der Daten auf dem Marktplatz in eigenen Analysen und Datenprodukten einen Mehrwert versprechen. Diese Skalierbarkeit ist jedoch, im Vergleich zu Unternehmen, die direkt Millionen Endkundinnen und -kunden sowie Haushalte ansprechen können, auf einen verhältnismäßig kleinen Kundenstamm von Business-Kundschaft begrenzt. Daher müssen die Datenmarktplatzbetreibende ihre Kosten für die Entwicklung der Plattform und zur Akquise der Kundschaft (Customer Acquisition Costs) auf diese wenigen Kundinnen und Kunden umlegen und versuchen, trotzdem ein qualitativ attraktives Angebot bereitstellen zu können.

Parallel zu diesem Paradigma hat sich ein ganzer Wirtschaftszweig um das Paradigma des Multi-Sided-Markets¹⁹ und der Plattformökonomie²⁰ entwickelt. Ein sehr frühes Beispiel für Deutschland sind die Unternehmen Interhyp AG und die Ende der 1990er Jahre gegründete Hypoport AG wurde, die heute wichtige Prozesse der Immobilienfinanzierung zwischen Banken und den Käuferinnen und Käufern vermittelt. Das Unternehmen wurde durch das Plattformgeschäft eines der ersten „Unicorns“ in Deutschland. Dabei sammelt das Unternehmen zahlreiche Daten über Kern-Entitäten wie die Käuferin bzw. den Käufer, die Bank bzw. das Kreditprodukt oder die Immobilie. Anhand dieser Daten erfolgt dann ein Matching der Käuferinnen bzw. Käufer und der Wunschimmobilie mit Angeboten der Kreditfinanzierung und damit ein potenzielles Angebot an die Kundin bzw. den Kunden. Nimmt die Kundin bzw. der Kunde das Angebot wahr, werden die Profile der Kern-Entitäten mit den Daten aus der erfolgten Transaktion erweitert. Dadurch kann das Matching bei einer ähnlichen Kundschaft bzw. Immobilie noch genauer mögliche Kreditinstrumente zuordnen. Die Servicequalität erhöht sich und damit oft auch die Kundenzufriedenheit und die Weiterempfehlungsquote. Dadurch kommen neue Kundinnen bzw. Kunden und weitere transaktionale Daten können das Matching noch weiter verbessern.

Noch bekannter ist Amazon Inc., das ab 2004 Anbietende von Büchern an eine Kundschaft vermittelte und den Prozess des Einkaufs, „Schaufensterbummel“, Kaufen und Liefern, in einem Feedback-Loop implementierte. Zunächst erfolgte der Feedback-Loop auf einem schwachen Signal, nämlich wenn eine Kundin bzw. ein Kunde für mehrere Bücher Interesse zeigte, diese also beispielsweise innerhalb ei-

¹⁹ Vgl. Gassmann et al. 2018.

²⁰ Vgl. Cusumano et al. 2019.

ner Session anschaute. Auf Basis dieses Signals wurde eine Recommendation-Engine angewandt: Ähnliche Kundinnen bzw. Kunden erhielten ebenfalls die Bücher ihrer Peer-Gruppe angezeigt. Erste Versuche zeigten, dass dieser „Proxy“ für das Kundeninteresse zu schwach war. Daher wurden im zweiten Versuch nur noch Bücher für die Recommendation-Engine als Trainingsdaten verwendet, die tatsächlich auch zusammen gekauft wurden. Dieser Ansatz funktionierte deutlich besser. Der Feedback-Loop, das Lernen aus dem Verhalten der Kundschaft aus der Vergangenheit, und die Vereinfachung des Bucheinkaufs führten zu einem hohen Anstieg der Kundenzufriedenheit. Dadurch benutzten immer mehr Käuferinnen und Käufer die Empfehlungsseite von Amazon. Das führte zu mehr Varianz in den Kundenprofilen und insgesamt zu mehr Kundschaft in den Kern-Entitäten – oder Profilen – wodurch die Recommendation-Engine wiederum bessere Empfehlungen geben konnte. Ein Feedback-Loop war entstanden, bessere Daten gaben bessere Analysen, diese ermöglichten bessere Datenprodukte und diese zogen mehr Kundschaft an, wodurch wiederum zusätzliche komplementäre Daten dem Unternehmen zu Kundinnen bzw. Kunden zur Verfügung standen. Letzten Endes konnten auch die Entwicklungskosten der Plattform und der Feedback-Loop auf hunderte von Millionen von Transaktionen umgelegt werden und somit standen insgesamt mehr Gelder für die Weiterentwicklung der Plattform zur Verfügung.

3.2 Wiederverwendung von Profilen und Plattformtechnologie in weiteren Datenprodukten

Amazon sammelte schließlich in zahlreichen anderen Prozessen mit einem Feedback-Loop weitere Daten zu Profilen der Anbietenden von Büchern, z. B. zur Lieferverfügbarkeit, Pünktlichkeit, Qualität der Lieferung etc., und zu anderen Profilen, zunächst Bücher (wer kauft welches Genre, welche Qualität, Preissegmente etc.), aber auch zu anderen Produkten, wie Elektronik, Computer, Videos, Musik und schließlich sogar zu Lebensmitteln. Diese neuen Profile verwendete Amazon dann zunächst zur Optimierung der Qualität in weiteren Prozessen, wie dem Anbieten von Bezahlung auf Kredit, der Bewertung der Lieferantenbeziehungen und -ketten, der proaktiven Lagerhaltung, dem Schalten von Werbung und Marketinginstrumenten bis hin zur Ausrichtung der Produktion aufkommende Trends. Über die letzten zwei Dekaden konnte Amazon so in immer mehr Prozesse unseres privaten und geschäftlichen Lebens attraktive, personalisierte Datenprodukte auf seiner Plattform anbieten. Dabei konnte das Unternehmen die Kosten für die Entwicklung eines Datenproduktes durch Wiederverwendung existierender Profile senken und zahlreiche Feedback-Loops etablieren, die immer weiter die zentralen Profile der Kern-Entitäten in ihrem Multi-Sided-Markt verbessern.

Dadurch entstand ein Zyklus: Einerseits konnte das Unternehmen schnell und kosteneffektiv neue Datenprodukte durch Wiederverwendung von Profilen, skaliert

barer Plattformtechnologie sowie genormten Schnittstellen zu Datenprodukten und Profilen, Micro-Services, entwickeln. Auf der anderen Seite waren die Produkte sehr kundenfreundlich, schließlich gab es ja bereits viel Wissen über die Kundschaft oder zumindest ihre Peer-Gruppe aus anderen Datenprodukten, das immer öfter wiederverwendet und angereichert wurde. Dadurch konnten immer mehr Kundinnen und Kunden gewonnen werden und jede Kundin bzw. jeder Kunde zahlte einen kleinen Anteil mit jeder Transaktion an den Entwicklungskosten der Datenplattform.

3.3 Deep Learning zur Repräsentation von Geschäftsentitäten

Die technischen Möglichkeiten des Umgangs mit großen Datenmengen und das tiefe Lernen auf CPU bzw. später GPU-Clustern hat zusätzlich die Entwicklung von Datenprodukten in der Plattformökonomie vereinfacht. In den 2000ern wurde das Profil einer Kundin bzw. eines Kunden, eines Lieferanten, einer Werbung oder eines Mitbewerbers bzw. einer Mitbewerberin oft noch in einigen wenigen bis zu hundert Spalten einer Tabelle – oder Attributen eines Knowledge Graphen – repräsentiert. Dieser Ansatz resultierte aus der Möglichkeit, diese Profile zunächst ggf. aus einem Data Warehouse, aus Datenbanken oder Log-Dateien von Webservern wiederzuverwenden.

Die Daten wurden aufgrund von Compliance- und Reporting-Anforderungen schon vorgehalten und sollten jetzt für einen weiteren Zweck wiederverwendet werden. Die damaligen Lernverfahren benutzen oft Techniken des manuellen Feature-Engineerings, um aus diesen Tabellen die richtigen Merkmale für die Matching-Funktion des Datenproduktes abzuleiten. Diese Methodik ändert sich seit Beginn der 2010er Jahre massiv durch die Anwendung von Deep-Learning-Verfahren. Deep Learning erlaubt eine deutlich umfangreichere und vor allem latente Repräsentation der Profile durch Tausende bis Millionen von Neuronen in sehr tiefen Netzwerken; das aktuelle Netzwerk BERT²¹ zur Repräsentation von Sprache, z. B. in Kommentaren zu Produkten, erlaubt mehrere hundert Millionen von Neuronen. Dadurch konnten die Plattformbetreibenden noch mehr Daten aus noch mehr Transaktionen deutlich mehr Feinheiten und auch Anomalien unter den Profilen für Kundinnen und Kunden, Lieferanten etc. abbilden. Profile auf Basis von Deep Learning können aufgrund dieser umfangreichen Parameter daher noch stärker die Feinheiten der unterschiedlichen Wünsche der Kundschaft für das Matching berücksichtigen. Dieser Schritt von einer symbolischen Repräsentation von Kunden, Profilen oder Objekten, mit hunderten von Merkmalen zu einer eher latenten Repräsentation mit Millionen von Parametern findet derzeit überall statt. Er erlaubt viel feingranularer auch

²¹ Vgl. Devlin et al. 2019.

die Anomalien und Nischen über ein Merkmal oder eine Vielzahl von Merkmalskombinationen abzubilden. Genau diese feinen Nuancen werten ggf. ein Datenprofil auf und machen es umso wertvoller (siehe auch der Rechner der Financial Times zum Wert des eigenen Datenprofils im Abschnitt Preisbildung).

Auf der anderen Seite ermöglicht Deep Learning oft auch die multimodale Datenintegration. Beispielsweise hat das Unternehmen Zalando in der FashionDNA²² bildliche, tabellarische und textuelle Attribute seiner Produkte in einem 200-dimensionalen Vektorraum abgebildet. Damit kann das Matching zwischen Suchanfrage und Produktkatalog nun erstmals auch komplexe Anfragen, wie z. B. „Hippie-Bluse mit Troddeln“ matchen, deren Worte nur von wenigen Kundinnen bzw. Kunden stammen und damit auch nicht in den Metadaten der Produkte der Hersteller abgebildet sind. Das Verfahren lernt vielmehr aus dem Archiv/der Historie Assoziationen der Suchbegriffe mit dem Klickverhalten der Kundschaft und kann, als Nebeneffekt des Matchings, sogar die Profile der Kundinnen und Kunden sowie der Produkte weiter augmentieren.

In der Zwischenzeit dominieren die Unternehmen der Plattformökonomie weltweit die „klassischen“ Unternehmen. Die Top-5 Unternehmen der Plattform-Ökonomie in den Vereinigten Staaten (Apple, Amazon, Facebook, Microsoft und Alphabet) haben zusammen einen Marktwert von mehr als 3 000 Milliarden Euro,²³ das ist ungefähr die Größenordnung des Bruttoinlandsproduktes von Deutschland. Analog gilt das für die größten Unternehmen in Asien, wie Alibaba, Tencent, Samsung, Didi Chuxing, Baidu, oder Meituan.

Auch das oben bereits behandelte Thema der Preisbildung wird im Kontext von Datenmarktplätzen, die maschinelles Lernen einsetzen, neu diskutiert. Chen et al. beschreiben z. B. eine Modell-basierte Preisfindung, bei der nicht mehr die Daten bepreist werden, sondern die Instanzen des verwendeten Machine-Learning-Modells.²⁴ Ein weiterer Gegenstand der Preisbildung sind Trainingdaten für KI-Anwendungen bzw. für Anwendungen des maschinellen Lernens²⁵.

²² Vgl. Bracher et al. 2019.

²³ Vgl. Holger Schmidt, <https://www.netzoekonom.de/plattform-index/>.

²⁴ Vgl. Chen et al. 2019.

²⁵ Vgl. Agarwal et al. 2019.

Abb. 4: Feedback-Loops erzeugen neue transaktionale Daten für das maschinelle Lernen. Dadurch werden Kundinnen-/Kunden- und Lieferantenmodelle, Wettbewerberradar etc. des Plattformanbieters verbessert. Diese verbesserten Dienstleistungen führen zur besseren Kundenzufriedenheit. Zufriedene Kundinnen und Kunden geben Weiterempfehlungen, wodurch noch mehr Menschen die Plattform nutzen und zusätzliche transaktionale Daten in das maschinelle Lernen der Dienste der Plattformbetreibenden fließen. Diese verbesserte Servicequalität schafft dadurch Netzwerkeffekte, die die Plattform erreicht neue Kundinnen und Kunden und skaliert. Rückkopplungsschleifen verbessern existierende AI-(Daten)-Produkte bzw. ermöglicht die Schaffung neuer Datenprodukte.

3 Fazit und Ausblick

Datenmarktplätze greifen den Gedanken auf, Daten auf zwei-(oder mehr-)seitigen Märkten zu handeln. Dabei übernimmt der oder die Besizende des Marktplatzes die Funktionen der Präsentation, des Matchings, der Preisgestaltung und des Vertriebs. Datenmarktplätze existieren für einige Domänen schon sehr lange, in den frühen 2010ern wurde dieses Konzept vertieft wissenschaftlich untersucht. Datenmarktplätze sind ein typisches Business-to-Business-Produkt und imitieren in Teilen Technologien des ETL bzw. in der Bereitstellung von Datenpipelines. Allerdings werden auch bereits Überlegungen angestellt, dass Daten in Zukunft zum vollständigen Geld- bzw. Preisersatz werden könnten.²⁶

Die Plattformökonomie hat in den letzten zwei Dekaden eine deutlich bessere Möglichkeit der Skalierung des Geschäftsmodells erreicht als die Datenmarktplätze. Wichtige Elemente sind Businessmodelle auf Basis von Multi-sided Markets, Profilen von Kern-Entitäten oder Matching-Funktionen für die Optimierung zentraler Geschäftsmetriken (Key Performance Indicators) wichtiger Businessprozesse. Ebenfalls dazu gehören Feedback-Loops, die Profile weiter anreichern, Matching-Funktionen

²⁶ Vgl. Ramge und Mayer-Schönberger 2017.

und Datenprodukte verbessern und so neue zufriedene Kundinnen und Kunden gewinnen können. Durch Wiederverwendung der Profile in zahlreichen Geschäftsprozessen und Datenprodukten, standardisierten Schnittstellen (Micro-Services) und dem Einsatz von Deep Learning konnten die Kosten für die Entwicklung der Plattform und der Datenprodukte weiter gesenkt werden. Auf der anderen Seite konnten durch den Neugewinn der Kundinnen und Kunden nicht nur reichere Profile erzielt werden, vielmehr können auch die Kosten für die Entwicklung der Plattform auf mehrere Kundinnen und Kunden umgelegt werden. Aktuell dominieren amerikanische und asiatische Unternehmen die Plattformökonomie,²⁷ insbesondere in zahlreichen, sogenannten Winner-takes-most- oder Business-to-Customer-Märkten. Im Business-to-Business-Geschäft ist die Plattformökonomie noch weniger vertreten und hier bieten sich Chancen für die europäische Industrie.

Neue Trends zeigen sich bereits in dem Zusammenschluss führender Plattform-Monopole und dem Versuch des Austauschs von Profilen über verschiedene komplementäre Plattformen hinweg. Ein Beispiel ist die Einführung der Crypto-Plattform von Libra von Facebook. Über eine in der Schweiz registrierte Legal-Entity, die Libra Association,²⁸ haben sich erstmals mehrere Vertreter der Plattformökonomie geeinigt, Transaktionen in deren Plattformen und den Bezahlverkehr gemeinsam zu regeln. Diese Unternehmen decken weltweit größtenteils komplementäre Märkte ab bzw. formen in den Märkten ein Monopol bzw. Oligopol. Dazu gehören Facebook (Werbung), Music Streaming (Spotify), Online-Zahlungsverkehr (Paypal, Visa, Mastercard, Stripe), Telco (Vodafone), Online-Marktplätze (eBay), Fahrdienstvermittler (Uber, Lyft) sowie Finanzdienste (Coinbase und Kiva).²⁹ Technologisch basiert das Konzept für den Austausch der Profile und der Transaktionen auf sog. Permissioned Blockchains. Diese kryptographisch geschützte, über Unternehmensgrenzen hinweg verteilte Datenstruktur erlaubt nur ausgesuchten, vorher registrierten Unternehmen das Lesen und Schreiben von Transaktionen und eignet sich besonders gut für Oligopole. Auch hier sind spezielle Datenmarktplätze in der Entwicklung.³⁰

²⁷ Vgl. „Künstliche Intelligenz: Potenzial und nachhaltige Veränderung der Wirtschaft in Deutschland“ Studie: Arthur D Little, Vodafone und Eco 2019.

²⁸ S. <https://www.diem.com/en-us/association/>.

²⁹ S.a. Facebook.com, Spotify.com, PayPal.com, Visa.com, Mastercard.com, Stripe.com, Vodafone.com, eBay.com, uber.com, Lyft.com, Coinbase.com, Kiva.com.

³⁰ Vgl. Travizano et al. 2018.

Literatur

Letztes Abrufdatum der Internet-Dokumente ist der 15.11.2020.

- Agarwal, Anish, Munther Dahleh und Tuhin Sarkar. 2019. „A Marketplace for Data: An Algorithmic Solution.“ In *EC '19: Proceedings of the 2019 ACM Conference on Economics and Computation/June 2019*, hg. v. Association for Computing Machinery, New York (NY), USA, 701–726. New York: ACM. doi:10.1145/3328526.3329589.
- Banerjee, Prabal und Sushmita Rujy. 2019. „Blockchain Enabled Data Marketplace – Design and Challenges.“ <https://arxiv.org/pdf/1811.11462.pdf>.
- Bracher, Christian, Sebastian Heinz und Roland Vollgraf. 2016. „Fashion DNA: Merging Content and Sales Data for Recommendation and Article Mapping“ Computing Research Repository (CoRR) abs/1609.02489.
- Chen, Lingjiao, Paraschos Koutris und Arun Kumar. 2019. „Towards Model-based Pricing for Machine Learning in a Data Marketplace.“ *Proc. ACM SIGMOD Int. Conf. on Management of Data, Amsterdam, Netherlands*: 1535–1552.
- Cusumano, Michael, Annabelle Gawer und David B. Yoffie. 2019. *The Business of Platforms: Strategy in the Age of Digital Competition, Innovation, and Power*. New York (NY), USA: Harper Business.
- Devlin, Jacob, Ming-Wei Chang, Kenton Lee und Kristina Toutanova. 2019. „BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.“ *NAACL-HLT* (1): 4171–4186.
- Gassmann, Oliver, Karolin Frankenberger und Michaela Csik. 2018. *Der St. Galler Business Model Navigator*. München: Hanser Verlag.
- Goldhammer, Andre und Andre Wiegand. 2017. *Ökonomischer Wert von Verbraucherdaten für Adress- und Datenhändler. Studie im Auftrag des Bundesministeriums der Justiz und für Verbraucherschutz*. Berlin: Goldmedia.
- Koutris, Paraschos, Prasang Upadhyaya, Magdalena Balazinska, Bill Howe und Dan Suciu. „Query-Based Data Pricing.“ *Journal of the ACM* 62 (5): Article 43.
- Lange, Julia, Florian Stahl und Gottfried Vossen. 2018. „Datenmarktplätze in verschiedenen Forschungsdisziplinen: Eine Übersicht.“ *Informatik-Spektrum* 41(3): 170–180.
- Martins, Denis, Jens Lechtenböcker und Gottfried Vossen. 2019. „Supporting Customers with Limited Budget in Data Marketplaces.“ *Proc. 6th IEEE Latin American Conference on Computational Intelligence (LA-CCI), Guayaquil, Ecuador*: 201–206.
- Meisel, Lukas und Markus Spiekermann. 2019. *Datenmarktplätze – Plattformen für Datenaustausch und Datenmonetarisierung in der Data Economy*. Techn. Report, Fraunhofer Institute for Software and Systems Engineering (ISST), Februar 2019.
- Muschalle, Alexander, Florian Stahl, Alexander Löser und Gottfried Vossen. 2013. „Pricing Approaches for Data Markets.“ In *BIRTE 2012 (Proc. 6th International Workshop on Business Intelligence for the Real Time Enterprise 2012, Istanbul, Turkey)*, hg. v. M. Castellanos, U. Dayal, E. Rundensteiner, 129–144. Heidelberg: Springer LNBP 154.
- Ramge, Thomas und Victor Mayer-Schönberger. 2017. *Das Digital: Markt, Wertschöpfung und Gerechtigkeit im Datenkapitalismus*. Berlin: Econ.
- Schomm, Fabian, Florian Stahl und Gottfried Vossen. 2013. „Marketplaces for Data: An Initial Survey.“ *ACM SIGMOD Record* 42 (1): 15–26.
- Stahl, Florian, Alexander Löser und Gottfried Vossen. 2015. „Preismodelle für Datenmarktplätze.“ *Informatik-Spektrum* 38(2): 133–141.
- Stahl, Florian, Fabian Schomm, Lara Vomfell und Gottfried Vossen. 2017. „Marketplaces for Digital Data: Quo Vadis?“ *Computer and Information Science* 10 (4): 22–37.

- Stahl, Florian, Fabian Schomm und Gottfried Vossen. 2014. „The Data Marketplace Survey Resited.“ *Proc. 11th International Baltic Conference on Databases and Information Systems (Baltic DB&IS Conference)*. Tallinn, Estonia: 135–146.
- Stahl, Florian, Fabian Schomm, Gottfried Vossen und Lara Vomfell. 2016. „A Classification Framework for Data Marketplaces.“ *Vietnam Journal of Computer Science* 3 (3): 137–143.
- Stahl, Florian und Gottfried Vossen. 2016. „Data Quality Scores for Pricing on Data Marketplaces.“ *Proc. 8th International Conference on Intelligent Information and Database Systems (ACIIDS)*, Da Nang, Vietnam, 215–224. ORT: Springer LNAI 9621.
- Stahl, Florian und Gottfried Vossen. 2017. „Name Your Own Price on Data Marketplaces.“ *Informatica – An International Journal* 28 (1): 1–26.
- Tang, Ruiming, Antoine Amarilli, Pierre Senellart und Stephane Bressan. 2014. „Get a Sample for a Discount – Sampling-Based XML Data Pricing.“ *Proc. Int. Conf. on Database and Expert Systems Applications (DEXA)* Springer LNCS 8644, 20–34.
- Travisano, Matias, Carlos Sarraute, Gustavo Aizenman und Martin Minnoni. 2018. „Wibson: A Decentralized Data Marketplace.“ *Proc. SIGBPS Workshop on Blockchain and Smart Contract*, San Francisco, 2018.