

Preismodelle für Datenmarktplätze

Florian Stahl · Alexander Löser
Gottfried Vossen

Einleitung

Im Zeitalter der umfassenden und alle Bereiche des täglichen Lebens erreichenden Digitalisierung entstehen derzeit 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 [10], 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 mittelständische Unternehmen und Fachabteilungen in Konzernen verfügen oft über spezifisches Wissen für eine Übersetzung von Daten in monetär verwertbare Informationen, bspw. Fachwissen im Bezug auf eine spezielle Kunden- oder Produktnische oder Beziehungen zu potenziellen Kunden, die bereit sind, für diese Informationen zu bezahlen. Zur Analyse der Daten sind IT-Infrastrukturen notwendig, die Web-Daten sammeln und mit Informationen über ihre Herkunft, ihren Inhalt und ihre sprachliche Struktur anreichern. In der Regel erfordern derartige Unterfangen erhebliche Infrastrukturinvestitionen und erzeugen erhebliche weitere Kosten für die Aktualisierung von Daten, welche sich insbesondere Mittelständler und Fachabteilungen oft nicht leisten können.

Datenverkäufer, Data-Warehouse-Entwickler sowie Entwickler von domänenspezifischen Mining-Algorithmen bieten heute ihre Verfahren auch als Plattform-, Software- und Data-as-a-Service-Modell

auf derartigen Datenmarktplätzen an. Der zentrale Einstiegspunkt eines solchen Marktplatzes erleichtert den Zugang zu diesen Dienstleistungen und Daten. Zusätzlich fungiert er als kunden- und anbieterübergreifende Datenintegrationsplattform, insbesondere für das kollektive Speichern, Analysieren und Wiederverwenden von Daten. Beispielsweise kann ein Anwendungsprogrammierer für spezielle Marktsegmente auf der Basis von offenen und kommerziellen Daten auf dem Marktplatz eigene Tools für die Marktforschung anbieten. Damit ermöglichen Datenmarktplätze völlig neue Geschäftsmodelle mit Informationen und Analysetools als elektronisch handelbaren Gütern.

In diesem Zusammenhang ergeben sich unter anderem folgende Fragen: Wer agiert auf und profitiert von einem Datenmarktplatz? Welche Preismodelle wählen Anbieter von Daten und datennahen Dienstleistungen und warum? Welche technischen Herausforderungen ergeben sich aus dem Handeln von Daten und den zugehörigen Diensten auf Datenmärkten? Dieses Papier stellt eine empirische Studie mit zwölf Top-Level-Managern und Experten aus europäischen und nordamerikanischen Unternehmen vor, die Datenmarktplätze betreiben bzw.

DOI 10.1007/s00287-013-0751-7
© Springer-Verlag Berlin Heidelberg 2013

Florian Stahl · Gottfried Vossen
WWU Münster, ERCIS
Leonardo-Campus 3
48149 Münster
E-Mail: flst@ercis.de, g.v@wwu.de

Alexander Löser
Beuth Hochschule für Technik Berlin, DATEXIS
Luxemburger Straße 10
13353 Berlin
E-Mail: aloeser@beuth-hochschule.de

Zusammenfassung

Eine zunehmende Zahl von Anbietern nutzt das Cloud-Computing-Paradigma für einen Handel mit Daten und analytischen Dienstleistungen. In dieser qualitativen Studie präsentieren wir die Ergebnisse aus Interviews mit zwölf etablierten Anbietern. Unsere Ergebnisse zeigen insbesondere eine große Unsicherheit bezüglich der Preissetzung und Preismodellwahl. Ferner erlauben sie eine Abstraktion der betrachteten Marktplätze auf ein einheitliches Schema mit sieben Akteuren sowie sechs atomaren und zwei hybriden Preisstrategien zu abstrahieren. Darüber hinaus bietet diese Papier erstmals eine strukturierte Entscheidungshilfe für die Wahl eines geeigneten Preismodells für Datenmarktplätze und legt somit den Grundstein für eine algorithmische Unterstützung bei Preismodellwahl und Preisfindung.

datenbezogenen Dienstleistungen anbieten. Dazu haben wir zunächst sieben Interviews durchgeführt und durch einen Snowballing-Mechanismus [2] eine zweite Serie mit fünf Interviewpartnern zusammengestellt. Unsere Interviewpartner kommen aus unterschiedlichen Teilgebieten der datenbezogenen Produkte und Dienstleistungen, insbesondere Social Media Monitoring, Text-Enrichment, Beratung und Datenmarktplätze.

Methodisch sind wir einem 7-Stufen-Ansatz nach [4] mit semistrukturierten Telefoninterviews gefolgt. So konnten wir gleichzeitig Flexibilität während der Interviews und Vergleichbarkeit der Ergebnisse erreichen. Während der durchschnittlich 66 min langen Interviews haben wir die Themenbereiche Erfahrung der Teilnehmer, Produkte und Geschäftsmodelle thematisiert und schließlich besprochen, wie ein idealer Datenmarktplatz aussehen sollte.

Teile der vorliegenden Arbeit wurden bereits in [5] veröffentlicht; wir stellen hier Erkenntnisse aus dem Gesamtprozess vor und gehen dabei insbesondere auf Preismodelle und Preismodellfindung für Daten und datenbezogene Dienstleistungen ein und fokussieren auf Herausforderungen, die (un-)mittelbar mit der Berechnung von Preisen zu tun haben. Hintergrund ist das Ziel, aus den hier beschriebenen ökonomischen und organisatori-

schen Erkenntnissen Preisfindungsmethoden und -algorithmen für Datenmarktplätze zu entwickeln und zu implementieren.

Die Arbeit ist wie folgt strukturiert: Zunächst stellen wir im Abschnitt „Datenmarktplätze“ das Konzept *Datenmarktplatz* im Detail vor und identifizieren seine (sieben) Akteure. In Abschnitt „Preismodelle und Einflussfaktoren“ beleuchten wir Preismodelle für Marktplatzbetreiber. Verwandte Arbeiten diskutieren wir in Abschnitt „Verwandte Arbeiten“. Schließlich stellen wir in Abschnitt „Zusammenfassung und Ausblick“ neue Forschungsfragestellungen vor, die vor allem aus dem ökonomischen Verhalten der Marktplatzteilnehmer resultieren und legen unser weiteres Vorgehen dar.

Datenmarktplätze

In diesem Abschnitt erläutern wir die Kernfunktionalitäten eines Datenmarktplatzes und sieben charakteristische Nutzergruppen (Akteure), basierend auf den Erkenntnissen aus unseren Interviews.

Infrastruktur

Ein Datenmarktplatz stellt potenziellen Kunden sowohl Daten als auch datennahe Dienstleistungen bereit. Oft können Kunden diese ohne einen Datenmarktplatz nicht oder nur zu unverhältnismäßig hohen Kosten beziehen. Die Datenmarktplätze unserer Interviewpartner integrieren Daten von öffentlichen Quellen (Web, UNO, Regierungen), nichtöffentlichen Quellen (Börsendaten) und In-House-Quellen (ERP-Systemen). Häufig stellt der Datenmarktplatz auch Infrastruktur zur Verarbeitung der Daten bereit. Wie in Abb. 1 zu sehen, basiert ein Datenmarktplatz auf einer skalierbaren Infrastruktur zur Verarbeitung größerer Datenmengen, beispielsweise Ausschnitte aus dem Web. Außerdem existiert eine Schnittstelle zum Bereitstellen von komplementären User Defined Functions (UDFs) für das Data-Mining, Data-Cleansing und für die Datenintegration. Weiterhin stellt der Datenmarktplatz Komponenten für Datenabruf, Handel und Abrechnung der Basisinfrastruktur sowie der UDFs zur Verfügung.

Akteure und Ökosystem

Im Rahmen unserer Interviews konnten wir sieben Akteure identifizieren, die von einem Datenmarkt-

Abstract

An increasing number of providers is using the cloud computing paradigm for trading data and providing analytical services. In this qualitative study we present the results from interviews with twelve well-established data providers. Our results in particular indicate that there is a great uncertainty regarding the choice of a pricing model and deriving appropriate prices. In addition, they allow to abstract the marketplaces considered into a unified schema with seven actors as well as six atomic and two hybrid pricing strategies. Additionally, for the first time, this study provides structured decision guidance regarding the choice of a suitable pricing model for data marketplaces. Thus, this paper lays the foundation for algorithmic support for choosing pricing models as well as support for pricing data and data-related goods.

platz profitieren (siehe Abb. 1). Diese lassen sich in direkte Profiteure (Datenanbieter, Analysten, Anwendungsentwickler, Algorithmenentwickler, Datenmarktplatzbetreiber) und indirekte Profiteure (Berater und Zertifizierungsagenturen) unterteilen:

- **Datenanbieter.** Unsere Interviewpartner unterscheiden Datenanbieter in kommerzielle und nichtkommerzielle Anbieter. In die erste Gruppe fallen „klassische“ Datenanbieter, wie bspw. Reuters oder Bloomberg, aber auch Suchma-

schinenbetreiber wie Google oder Bing sowie Anbieter von Linked Data und Web-Forenbetreiber, die ihre Inhalte monetarisieren wollen. Nichtkommerzielle Datenanbieter sind vor allem Regierungen und Behörden, wie bspw. die UNO, die ihre Daten kostenlos anbieten wollen oder per Gesetz müssen.

In der Regel nutzen Datenanbieter Datenmarktplätze, um ihre Daten zu speichern und zu bewerben und sie so einem größeren Publikum zugänglich zu machen, was meist günstiger als der Eigenvertrieb ist. Daher fließen Daten vom Anbieter zum Marktplatz und die Datenanbieter bekommen Geldzuflüsse in Form von Umsatzbeteiligungen.

- **Analyst/Professional.** Typischerweise besteht diese Gruppe aus Domainexperten wie Produktmanagern, Brokern, oder Businessanalysten. Vom Angebot des Datenmarktplatzes erhoffen sich Analysten durch die Fülle an zur Verfügung stehenden Daten und deren Kombinationsmöglichkeiten genauere Analyseergebnisse. Mitglieder dieser Gruppe stellen – je nach Expertise und technischen Möglichkeiten – strukturierte und textbasierte bzw. Ad-hoc-Anfragen an den Datenmarktplatz. Auf diese Art und Weise formalisieren sie ihr (häufig domänenspezifisches) Wissen über Datenintegrationsschritte. Der Marktplatzbetreiber profitiert direkt durch die Entgelte und indirekt über die Möglichkeit, diese Nischenwissen ggfs. in Form von UDFs Dritten zugänglich zu machen.
- **Anwendungsentwickler/-verkäufer.** Häufig sind Analysten nicht in der Lage, ihre Anforderungen

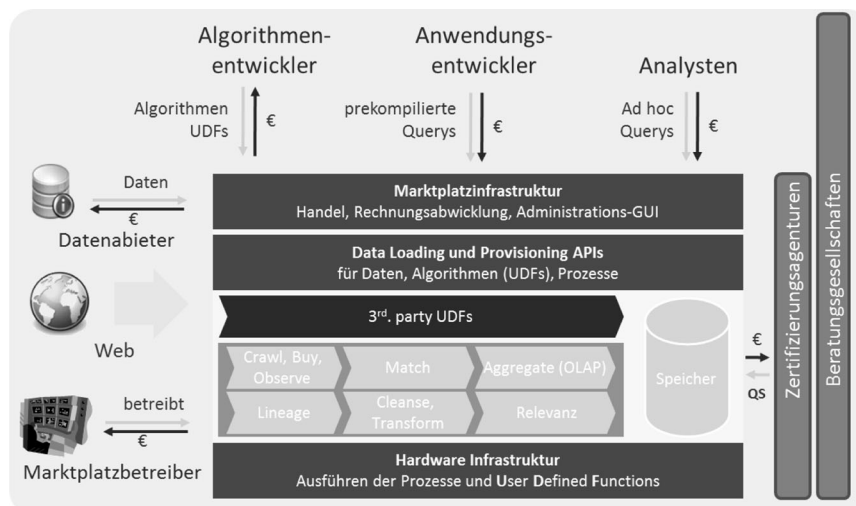


Abb. 1 Schematische Darstellung eines Datenmarktplatzes



Tabelle 1

Preiseinflussfaktoren		
Faktor	Ausprägungen	Preiseinfluss
Art der Daten	Statistisch-faktisch, aktuell	Statisch-faktische Daten sind i. d. R. günstiger
Umfang der Daten	Datensatz komplett, Ausschnitt	Komplette Datensätze sind i. d. R. teurer
Form der Daten	Strukturiert, unstrukturiert	Strukturierte Daten sind i. d. R. teurer
Datenqualität	–	Qualitative Daten (bswp. vollständig, integrierbar) sind i. d. R. teurer
Art der Nutzung	Eigennutzung, Vertrieb, beliebige	I. d. R. steigt der erzielbare Preis mit steigenden Rechten.
Nutzungsdauer	temporär, unbegrenzt	Temporäre Nutzung ist i. d. R. günstiger
Kostenträger	Nachfrager, Anbieter	Nachfrager akzeptieren i. d. R. höhere Preise
Marktform	Polypol, Oligopol, Monopol	I. d. R. lassen sich desto höhere Preise durchsetzen, je monopolistischer der Anbieter agieren kann
Empfundener Wert	–	Je nützlicher die Daten für den Kunden desto teurer sind sie i. d. R.
Exklusivität	–	Je exklusiver die Daten für den Kunden desto teurer sind sie i. d. R.

in einer formalisierten Sprache auszudrücken. So hat sich ein Markt für Anwendungsentwickler etabliert, welche Applikationen entwickeln, die Anfragen an Datenmarktplätze vereinfachen, in dem sie „Wissen“ über Datenschemata, Datenintegration und Quellenauswahl in vorkompilierte Anfrage übersetzen. Der Marktplatzbetreiber erhält Nutzungsentgelte für die bereitgestellten Daten, die Nutzung der Infrastruktur und die Vermittlung von möglichen Komponenten für die Datenanreicherung und -bereinigung. Typische Konsumer-Anwendungen sind Apps wie „Around Me“, die Geodaten verarbeiten, oder Finanzticker für Privatpersonen.

- **Entwickler von UDFs.** Die beiden vorgenannten Akteure, Analysten und Anwendungsentwickler müssen häufig große Datenmengen aus meist volatilen und heterogenen Datenquellen integrieren. Dafür werden UDFs für Data Mining, -Matching, -Cleansing, -Relevance Computation und -Lineage eingesetzt. Diese sind häufig auf eine bestimmte Domäne, Sprache, Datenquelle oder Funktionalität fokussiert. Ein Datenmarktplatz bietet UDF-Entwicklern die Möglichkeit, ihre Algorithmen als Black-Box-User-Defined-Function zu hinterlegen, sodass andere Akteure diese Algorithmen kaufen oder mieten können, um sie für ihre Zwecke einzusetzen. UDF-Entwickler

werden über Umsatzbeteiligungen abhängig von der Nutzung ihrer UDFs entlohnt.

- **Marktplatzbetreiber.** Der Betreiber ist das Wirtschaftssubjekt, das die Aufwendungen und das Risiko für den Betrieb der technischen Infrastruktur trägt, dafür aber die Entgelte der Marktteilnehmer vereinnahmt. Dabei sehen sich Betreiber mit einer Vielzahl an technischen, ethischen, rechtlichen und wirtschaftlichen Herausforderungen konfrontiert. Hier betrachten wir vornehmlich die wirtschaftlich-technischen Herausforderungen.
- **Berater.** Unsere Interviewpartner weisen häufig auf den Beratungsbedarf vornehmlich der Nutzer im Umfeld eines Marktplatzes hin. So sollen diese dabei unterstützt werden, Quellen auszuwählen oder eigene Produkte auf Basis der vorhandenen Daten zu entwickeln. Der Datenmarktplatz generiert also indirekt Kunden für die Beratungsgesellschaften. Allerdings profitiert auch der Marktplatzbetreiber, da so einerseits eine Nutzung sichergestellt wird und andererseits die Kunden eine technische Betreuung erfahren, für die der Betreiber nicht selbst aufkommen muss.
- **Zertifizierungsagenturen.** Die Aufgabe dieser Gruppe ist es, die Sicherheit der Plattform und die Qualität der Daten und Prozesse zu zertifizieren. Durch entsprechende Zertifikate kann die

Plattform das Vertrauen bei den Kunden erhöhen und Marktanteile gewinnen, wobei die Kosten der Zertifizierung i. d. R. an den Kunden weitergegeben werden.

Preismodelle und Einflussfaktoren

In diesem Abschnitt beschreiben wir Einflussfaktoren auf den Preis sowie beobachtete Preismodelle und geben eine Hilfestellung bei der Preismodellwahl. Zwar nennen wir zu den Preismodellen Beispiele für Marktplätze und Datenanbieter, die diese verwenden, doch kann hier kein umfassender Bericht über die Verteilung von Preismodellen gegeben werden. In [9] ist eine quantitative Studie des Marktes zu finden.

Einflussfaktoren auf den Preis für Daten

Aus den Interviews konnten wir drei grundlegende Gruppen von Einflussfaktoren ableiten, die den Preis für Daten und datennahe Dienstleistungen beeinflussen:

1. **Datenbezogene Einflussfaktoren.** Die Wertbarkeit von Daten hängt von verschiedenen Faktoren ab, die den Daten inhärent sind. Dazu zählen unter anderem die *Art*, die *Form* und der *Umfang* der Daten, aber auch die *Datenqualität*. Zu letzterer gehören insbesondere berechenbare Kriterien wie Vollständigkeit oder Dichte (vgl. [6]).
2. **Rechtebezogene Einflussfaktoren.** Der Wert von Daten ist auch davon abhängig, wie umfänglich die Nutzungs- und Wiederverwertungsrechte sind. Gängig sind Einschränkungen bezüglich der zeitlichen Nutzungsdauer oder der Weitergabe der Daten.
3. **Ökonomische Einflussfaktoren.** Die identifizierten ökonomischen Faktoren sind *Marktform*, *Kostenträger*, subjektives *Werteempfinden* und *Exklusivität* des Zugangs zu den Daten.

Tabelle 1 gibt einen Überblick über die genannten Preiseinflussfaktoren. Dabei stehen die datenbezogenen Faktoren oberhalb der ersten Linie, die rechtebezogenen oberhalb der zweiten Linie und die ökonomischen darunter.

Beobachtete Preismodelle

In den Interviews identifizierten wir sechs atomare und zwei hybride (zusammengesetzte) Preismodelle, die als nächstes beschrieben werden.

Atomare Preismodelle.

Kostenlos-Modell. Statistische Daten von Behörden¹ stehen oft kostenlos im Internet zur Verfügung. Diese können zwar nicht unbedingt monetarisiert werden, können einem Marktplatz jedoch helfen, Kunden zu gewinnen, was wiederum hilft, Anbieter von kommerziellen Daten zu gewinnen. Darüber hinaus können kostenlose Daten mit In-house- oder privaten Daten integriert werden und so neue Werte schaffen. Ein Spezialfall ist, dass manche Behörden oder Non-Profit-Organisationen ihre Daten veröffentlichen müssen. Für sie kann es ggfs. preiswerter sein, die Publikation auf den Marktplatz auszulagern. Beispielsweise bietet datamarket.com öffentlich verfügbare Daten kostenlos an.

Tauschgeschäft. Es kommt in der Praxis vor, dass ein Datenhändler Daten kostenlos zur Verfügung gestellt bekommt, wenn der Datenanbieter im Gegenzug die Daten in veredelter Form zurückbekommt. Dies funktioniert allerdings nur, wenn es neben den Datenlieferanten noch weitere Abnehmer für das Endprodukt gibt. Beispielsweise bieten Marktforschungsgesellschaften Warenhändlern Marktforschungsberichte kostenlos an, wenn sie im Gegenzug deren Absatzzahlen erhalten. Zusätzlich werden die Berichte an Hersteller verkauft und so Gewinne erwirtschaftet.

Festpreis/Subskription/Abonnement. Dieses Preismodell hat als einzigen Parameter die Nutzungsdauer, für die ein Nutzungsrecht eingeräumt wird. Bei unseren Gesprächspartnern beobachteten wir dieses Preismodell vor allem in Bezug auf Softwarelizenzen und Software-Hosting; aber auch Nachrichtendienste wie z. B. Bloomberg verkaufen ihre Informationen nach diesem Schema.

Paketpreise/Tiered Pricing. In diesem Modell werden dem Kunden unterschiedlich gestaffelte Pakete (bspw. API-Aufrufe, Anzahl Datenquellen) zu einem festen Preis und für eine bestimmte Nutzungsdauer angeboten. Im Unterschied zum Festpreis wird neben der Zeit auch die Menge festgelegt. In der Regel sind größere Pakete insgesamt teurer, aber relativ pro Einheit günstiger. Einige Interviewpartner bevorzugten dieses Modell anstelle des Festpreismodells aufgrund der Flexibilität und der Planungssicher-

¹ Z. B. <http://data.gov.uk/>.

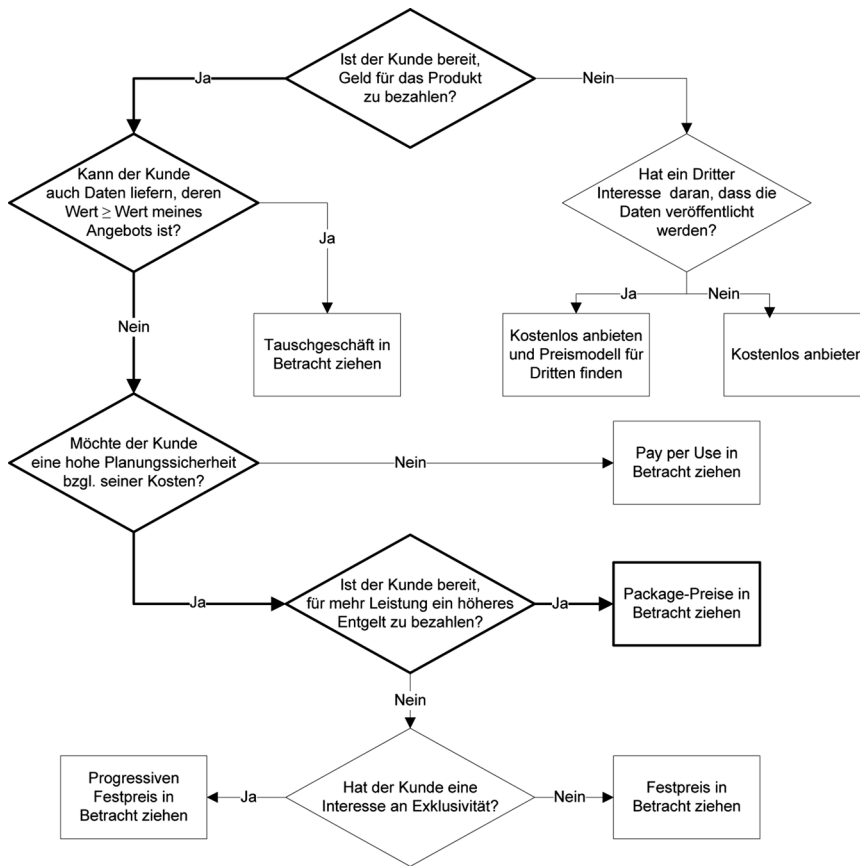


Abb. 2 Preismodell-entscheidungshilfe

heit. Konkret angewendet wird es von verschiedenen Anbietern auf dem Microsoft Azure Marketplace, bspw. von wt360.com².

Zahle-nur-was-Du-nutzt/Pay-per-Use. Dieses Preismodell setzt einen Preis für jede verbrauchte Einheit fest und bemisst den Endpreis anhand der verbrauchten Einheiten. Unsere Interviewpartner verwenden dieses Preismodell vorwiegend für die Abrechnung von Beratungszeit. Wir konnten dieses Preismodell nicht für die Abrechnung von API-Aufrufen oder dergleichen beobachten, da dieser volumenbasierte Ansatz für unsere Interviewpartner uninteressant wird, wenn die Grenzkosten der Datenbeschaffung und -verarbeitung gegen Null konvergieren. Sie erwarten aber, dass der Trend fallender Grenzkosten anhält und somit volumenbasierte Preise weiter an Attraktivität verlieren. Außerdem ist dieses Preismodell auch aufgrund der inhärenten Planungsunsicherheit (in Bezug auf

erwartete Absätze) nicht attraktiv. Häufig werden die Begriffe Pay-per-Use und Tiered Pricing synonym verwendet; ein Beispiel hierfür ist die Bayrische Staatsbibliothek³, die ihr Preismodell zwar Pay-per-Use nennt, nach unserer Definition jedoch Tiered Pricing betreibt. In einem angrenzenden Kontext kann man Pay-per-Use jedoch finden. So wird es bspw. von Amazon Web Services verwendet⁴, um Rechenzeiten in Rechnung zu stellen.

Progressiver Preis. Hier richtet sich der Preis eines Datensatzes nach dem Zeitpunkt des Erwerbs. Der Preis erhöht sich, je mehr Kunden eine nicht verkäufliche Lizenz zur Nutzung des Datensatzes erwerben. Dieses Preismodell wird vor allem angewandt, wenn die Verbreitung limitiert werden soll (u. a. auch bei künstlerischen Werken). Praktiziert wird dies bspw. von Dreamstime über sogenannte Preis-Level⁵.

² <http://datamarket.azure.com/dataset/weathertrends/worldwidehistoricalweatherdata>.

³ <http://www.bsb-muenchen.de/510.0.html>.

⁴ <http://aws.amazon.com/de/ec2/pricing/>.

⁵ <http://de.dreamstime.com/faqs-detail-1>.

Hybride Preismodelle.

Zweigeteilter Tarif/Two-Part-Tariff. Der zweigeteilte Tarif ist eine Kombination aus einer Grundgebühr (Festpreis, ggfs. progressiv) und entweder einem direkt (Pay-per-Use) oder indirekt (Paketpreis) verbrauchsabhängigem Anteil. Dieses Preismodell wurde in der Vergangenheit häufig von Strom- und Telekommunikationsanbietern verwendet und repräsentiert die hohen Investitionskosten dieser Branchen für Infrastrukturen. Wir beobachten bei unseren Interviewpartnern dieses Preismodell für Softwarelizenzen, in dem der Basispreis die Kosten für die Software deckt und Daten nutzungsabhängig berechnet werden. Häufig wird der feste Teil genutzt, um einen sicheren Deckungsbeitrag zu erzielen. Die variable Vergütung dient der Kundendiskriminierung nach Zahlungsbereitschaft. Als konkrete Beispiele sind hier Back-Up-Dienste wie JungleDisk⁶ und kineticD⁷ zu nennen, die neben einer Grundgebühr das Datenvolumen gigabytegenau abrechnen. Somit wird eine Grundgebühr mit einem nach unserer Definition echten Pay-per-Use kombiniert.

Freemium. Freemium ist ein Portmanteau aus Free, also kostenlos, und Premium. Die Idee ist, dass Benutzer grundlegende Produkte und Dienstleistungen kostenlos benutzen können, aber für den vollen Funktionsumfang ein Entgelt bezahlen müssen. Das Preismodell für die Premiumleistungen kann dabei jede der diskutierten Formen annehmen. Ein prominentes Beispiel ist Microsoft SQL Server. So gibt es unter dem Namen SQL Server Express eine kostenlose Variante, die allerdings Datenbanken mit maximal 10 GB unterstützt.⁸ Sind größere Datenbanken gewünscht, so muss man auf kostenpflichtige SQL-Server-Varianten zurückgreifen.

Auswahl des Preismodells

Unsere Interviewpartner stellen einheitlich den Bedarf nach Unterstützung bei der Preismodellwahl fest. Die Preismodelle *Kostenlos*, *Tauschgeschäft*, (*progressiver*) *Festpreis* und *Freemium* (je nach Ausgestaltung) haben keinen variablen Anteil und sind einfacher zu implementieren als Modelle mit variablem Anteil. Letztere sind nur unter der Voraussetzung möglich, dass die *Form des Angebots*,

also wie die Daten zugänglich gemacht werden (API, Web-Interface, etc.), eine Abrechnung der anfragespezifischen Nutzung ermöglicht.

Basierend auf den Kriterien aus Tab. 1 präsentieren wir in Abb. 2 eine Entscheidungshilfe für Preismodelle in Form eines Flow-Charts. Die zwei maßgeblichsten Kriterien für die Preismodelle *Kostenlos* bzw. *Tauschgeschäft* sind die Einschätzungen des Markts zum Wert der Daten. Für andere Preismodelle gelten folgenden Bedingungen: 1) Ist der Kunde bereit, für den Anschluss an den Dienst (ggf. inkl. Grundleistungen) zu bezahlen, kann ein zweigeteilter Preis in Betracht gezogen werden. 2) Wenn man Daten zu Testzwecken verschenken kann, kann Freemium als Modell in Betracht gezogen werden. Zu beachten ist, dass dieses Diagramm für jedes Produkt durchlaufen werden sollte.

Am häufigsten haben wir in den Interviews das Modell *Paketpreise* beobachtet, weshalb es in Abb. 2 durch fette Umrandungen und Pfeile kenntlich gemacht ist. Zusätzlich setzen einige Partner auf das Freemium-Modell: Zum einen können durch den Free-Aspekt neue Kunden gewonnen werden, die das Produkt risikolos testen und so sicher sein können, dass das Produkt ihren Anforderungen genügt. Zum anderen sind durch den Premiumaspekt weitere Umsätze wahrscheinlich. Gleichzeitig ermöglichen gestaffelte Paketpreise, als kostenpflichtige Komponente des Freemium-Modells, Kunden anhand der Nutzungshäufigkeit und -menge zu segmentieren und so besser deren Zahlungsbereitschaft abzuschöpfen. Darüber hinaus geben sie Planungssicherheit hinsichtlich des Umsatzes und der Nutzung.

Es mag so scheinen, dass diese Erkenntnisse wenig Relevanz für die Informatik haben. Um allerdings, wie es unser Ziel ist, in Zukunft Datenverkäufer und Datenmarktplatzbetreiber algorithmisch dabei unterstützen zu können, eine Preismodell oder gar einen konkreten Preis zu finden, ist dieses Wissen unerlässlich. Es dient also als wichtiges Hintergrundwissen für die bevorstehenden Arbeiten, auf die wir in Abschnitt „Zusammenfassung und Ausblick“ näher eingehen.

Verwandte Arbeiten

Grundsätzlich wurde das Thema Bepreisung von Informationsgütern bereits 1999 von Shapiro und Varian [11] behandelt. Unter anderem führen sie die Begriffe *Personalized Pricing* (kundenabhän-

⁶ <https://www.jungledisk.com/>.

⁷ <http://www.kineticd.com/>.

⁸ <http://www.microsoft.com/en-us/sqlserver/editions/2012-editions/express.aspx>.

gige Preise), Versioning (im Prinzip Paketpreise) und Group Pricing (unterschiedliche Preise je nach Nutzergruppe bspw. Geschäfts- und Privatkunden – im Prinzip Freemium oder Paketpreise) für die von Pigou bereits 1920 beschriebene Preisdiskriminierung 1. bis 3. Grades ein. Während unsere Studie zeigt, dass Versioning und Group Pricing bereits genutzt werden, ist das Personalized Pricing noch nicht implementiert. Dieses Papier soll als Grundlage dienen, dies usergenau für ein breites Spektrum an Informationsgütern zu ermöglichen.

Erst 2011 haben Balazinska et al. [1] Forschungsfelder zur Preissetzung von Daten identifiziert und eine Forschungsagenda erarbeitet. Dabei haben sie zwei Hauptherausforderungen bzgl. der Preissetzung identifiziert: Erstens müsse verstanden werden, wie der Wert von Daten auf Datenmarktplätzen festgelegt und modifiziert wird und welche Preismodelle und Services das Konzept des Datenmarktplatzes begünstigen. Die vorliegende Arbeit stellt mit dem Modell des Datenmarktplatzes und dem Entscheidungsbaum für die Modellauswahl einen wesentlichen Beitrag zu dieser Herausforderung dar. Die zweite Herausforderung bestehe darin, das Verhalten von Marktteilnehmern (und die zugrunde liegenden Regeln) zu verstehen. Diese zweite Herausforderung ordnen die Autoren den Wirtschaftswissenschaften zu und wollen die erste mithilfe der Datenbankcommunity lösen.

Die meisten Arbeiten im Bereich Preissetzung von Informationsgütern basieren auf der Annahme, dass Daten-Provider in der Lage sind, Preise monopolistisch zu setzen. Dabei werden vor allem die drei bekannten Strategien Festpreis, Pay-per-Use und Zweigeteilter Tarif betrachtet. In [14] wird dargelegt, dass, unter den Annahmen Monopol, homogene Kundenbasis und Grenz- und Überwachungskosten in Höhe von Null ein zweigeteiltes Preismodell ebenso gut ist wie ein Fixpreismodell. Wenn heterogener Konsum unterstellt wird, ist das zweigeteilte Modell am profitabelsten.

In [1] werden zwei Preisschemata identifiziert, die auf Datenmarktplätzen Anwendung finden: (a) ein Subskriptionsmodell mit einem Querylimit und (b) ein Preismodell, bei dem der Preis pro Datensatz (Tupel) festgelegt wird. Dabei werden die folgenden vier Unzulänglichkeiten der Preismodelle herausgearbeitet: 1.) Daten, die man einmal erworben hat, müssen entweder eigenständig gespeichert oder erneut erworben werden. 2.) Die Modelle ba-

sieren auf der Annahme, dass alle Datensätze gleich wertvoll sind 3.) Die Preismodelle erlauben Arbitrage. 4.) Die Anbieter von Daten erhalten keinerlei Hilfestellung bei der Preissetzung. Auf Basis unserer Interviews gehen wir in dieser Arbeit davon aus, dass die ersten drei nur aufgrund der vierten Unzulänglichkeit existieren. Unsere Entscheidungshilfe in Abschnitt „Preismodelle und Einflussfaktoren“ bietet erstmals eine Hilfestellung für die Preismodellauswahl.

Balazinska et al. [1] schlagen außerdem vor, Preise so zu modellieren, dass Preise an kleine Dateneinheiten (Zellen, Tupel etc.) geknüpft werden. Zusammen mit Regeln, wie sich diese Preise im Laufe der Verarbeitung einer Anfrage ändern, könne man endgültige Preise festlegen und damit ein Maximum an Flexibilität erreichen. Die Autoren schlagen weiter vor, Data Provenance zu nutzen, um dieses Problem zu adressieren. Das Konzept und eine Implementierung werden in [7] beschrieben. Kantere et al. [3] haben untersucht, wie Optimierungskosten den Benutzern über Anfragen in Rechnung gestellt werden können. In [13] wird ein spieltheoretischer Ansatz beschrieben, bei dem Benutzer aktiv auf eine Optimierung bieten. Eine Optimierung wird dann (und nur für die Bieter) installiert, wenn die Kosten durch die Summe der Gebote gedeckt werden.

Im Kontext des Cloud Computing wurde in [8] eine Methode vorgestellt, die Preise dynamisch generiert, um interaktive Verhandlungen oder Auktionen zu ermöglichen. Dabei schickt ein Interessent eine Anfrage an den Datenanbieter, dieser kalkuliert dynamisch den Preis und übermittelt ihn an den Interessenten, der nun akzeptieren oder ablehnen kann.

Zusammenfassung und Ausblick

In dieser Arbeit identifizieren wir durch eine qualitative Studie erstmals wesentliche Charakteristika weltweit agierender Datenmarktplätze. Wir konnten sieben Akteure identifizieren und beschreiben sechs atomare und zwei hybride Modelle für die Bepreisung von Daten und datennahen Dienstleistungen. Auf Basis unserer Analyse konnten wir erstmals eine domänenübergreifende Heuristik zur Auswahl eines Preismodells präsentieren.

Alle Interviewpartner sind einhellig der Meinung, dass Preise wertebasiert und somit abhängig von den Nachfragern erstellt werden müssen. Wie oben beschrieben ist dieser Wert hochgradig von

dem durch sie erzielten Informationsvorsprung abhängig. Oft fehlt jedoch einem Datenanbieter dieses Wissen und somit das Verständnis von dem „Wert“ seiner Daten für andere. Wir haben in dieser Arbeit drei Faktorgruppen (*ökonomisch, rechtlich, technisch*) identifiziert, die den potenziellen Informationsvorsprung und somit den Preis für Daten beeinflussen.

Unsere Interviewstichprobe ist offensichtlich zu klein für eine statistische Erhebung der diskriminierenden Preiskriterien. Daher ist weitere Forschung in zwei Richtungen notwendig: Um die *ökonomischen* und *rechtlichen* Kriterien zu ermitteln, planen wir eine Fragebogenstudie, die mittels Choice Modelling ausgewertet wird. So können wir Rückschlüsse auf die den Preis bestimmenden Faktoren ziehen. Allerdings werden solche Erkenntnisse hochgradig domänenspezifisch sein. Aus diesem Grunde werden wir uns zunächst der Datenqualität annehmen. Da diese den Daten inhärent ist, ist sie domänenunabhängig. Wie in Tab. 1 schon angeklungen, besteht Datenqualität aus einer Vielzahl von Faktoren, die es zunächst einzeln zu bestimmen und messbar zu machen gilt. Als vorläufiges Ergebnis, das es in weiteren Arbeiten zu bestätigen gilt, lässt sich festhalten, dass eine universelle Formel für den Nutzen von Daten lauten könnte:

$$U = \sum_{i=1}^n w_i e_i(x_1, \dots, x_n) + \sum_{j=1}^m v_j d_j(x_1, \dots, x_k)$$

Hierbei ist U der Nutzen eines Datums, der sowohl von 1 bis n ökonomischen und rechtlichen (e) sowie von 1 bis m datenspezifischen Funktionen (d) mit den Parametern x_1 bis x_n abhängt, jeweils mit spezifischen w bzw. v gewichtet. Aus dem Nutzen U bzw. der Verteilung des Nutzens U über die verschiedenen Nutzer hinweg lässt sich ein Produktpreis bzw. ein Preismodell bestimmen.

Somit konnte diese Studie erste Anhaltspunkte liefern, in welche Richtung Algorithmen entwickelt werden sollten, um letztendlich ein Preismodell zu empfehlen bzw. einen Preis zu berechnen. Diesbezüglich ist eine wichtige Aufgabe die Abstraktion vorhandenen Domänenwissens durch maschinelle Beobachtung des Verhalten von Akteuren auf Datenmarktplätzen. Aufgrund des Nischencharakters vieler Daten und Dienstleistungen erwarten wir eine Anwendung von aktiven Lernverfahren

bzw. von Lernverfahren, wie sie z. B. im Data-Mining [12] erfolgreich eingesetzt werden. Auch sind Preis-Absatz-Prognosen, wie sie in der Betriebswirtschaftslehre verwendet werden, auf einem Datenmarktplatz implementierbar, wenn verstanden ist, welche Faktoren den Preis beeinflussen und welche Zahlungsbereitschaft potenzielle Kunden haben. Für eine algorithmische Implementierung dieser Datenmarktplatzkomponenten konnten wir mit diesem Papier wichtige Grundlagen erarbeiten.

Acknowledgments

Wir danken den folgenden Interviewpartnern: Yasan Budak, CFO und Gründer, Vico Research & Consulting GmbH; Stefan Geissler, CEO und Gründer, TEMIS Deutschland GmbH; Holger Diewiger, CTO, Neofonie GmbH; Hjalmar Gislason, CEO und Gründer, datamarket.com; Ian Mulvany, VP Product Management, mendeley.com; Christoph Tempich, Head of Consulting, inovex GmbH; Phillip Mueller, VP Strategy Thomson Reuters; Bernhard Schweitzer, Direktor Research, SAP AG; Stefan Mueck, Executive Consultant, IBM Deutschland; Gil Elbaz, CEO und Gründer factual.com; Leigh Dodds, Open & Linked Data Consultant; und Thomas Ruf, selbständige Fachkraft im Bereich Marktforschung.

Literatur

1. Balazinska M, Howe B, Suciu D (2011) Data markets in the cloud: an opportunity for the database community. *PVLDB* 4(12):1482–1485
2. Bryman A, Bell E (2007) *Business Research Methods*. Oxford University Press, Oxford
3. Kantere V, Dash D, Gratsias G, Ailamaki A (2011) Predicting cost amortization for query services. In *SIGMOD Conference*
4. Kvale S, Brinkmann S (2008) *InterViews: Learning the Craft of Qualitative Research Interviewing*. Sage Publications, 325–336
5. Muschalle A, Stahl F, Löser A, Vossen G (2012) Pricing approaches for data markets. *BIRTE* 2012:129–144
6. Naumann F (2002) *Quality-Driven Query Answering for Integrated Information Systems*, vol 2261 of Lecture Notes in Computer Science, Springer
7. Balazinska M, Howe B, Koutris P, Upadhyaya P, Suciu D (2013) Toward practical query pricing with querymarket. *ACMSIGMOD SIGMOD Conference*: 613–624
8. Püschel T, Neumann D (2009) Management of Cloud Infrastructures: Policy-Based Revenue Optimization. In: *ICIS 2009 Proceedings*, paper 178
9. Schomm F, Stahl F, Vossen G (2013) Marketplaces for data: an initial survey. *SIGMOD Rec* 42(1):15–26
10. Löser A, Hueske F, Markl V (2009) Situational business intelligence. In: *Business Intelligence for the Real-Time Enterprise*. LNBI, Springer
11. Shapiro C, Varian HR (1999) *Information Rules: A Strategic Guide to the Network Economy*. Strategy/Technology/Harvard Business School Press
12. Tomanek K, Morik K (2011) Inspecting sample reusability for active learning. *Active Learning and Experimental Design @ AISTATS 2011*:169–181
13. Upadhyaya P, Balazinska M, Suciu D (2012) How to price shared optimizations in the cloud. *Proc VLDB Endow* 5(6):562–573
14. Wu SY, Banker RD (2010) Best pricing strategy for information services. *J AIS* 11(6):Article 1