

2 »Was mit Daten möglich ist«

Mit welchen Metaphern, Kategorien und Geschichten werden Daten beschrieben? Ein Blick auf populäre und feldspezifische Beschreibungen und Narrative sowie eine Konsultation sozialwissenschaftlicher Arbeiten zum Datenbegriff zeigen: Daten werden einerseits als universelles Mittel gefeiert, mit denen sich beliebige Zwecke erreichen lassen. Andererseits trifft diese rhetorische Übertreibung, dass mit Daten alles möglich sein soll, auf die unternehmerische Herausforderung, mit Daten etwas Konkretes möglich zu machen beziehungsweise herauszufinden, »was mit Daten möglich ist«.

Im zeitgenössischen Datenbegriff steckt eine Grundspannung: Auf der einen Seite gelten Daten als inhärent wertvoller Rohstoff, der für verschiedenste Anwendungen gebraucht werden kann. Auf der anderen Seite weisen Datenkritik – aber auch Datenpraktiker – darauf hin, dass Daten immer lokal situiert sind. Daten können niemals »roh« sein, weil sie immer auf etwas verweisen, das außerhalb von ihnen liegt. Diese Gleichzeitigkeit ist folgenreich und zieht sich durch die verschiedenen Inkarnationen von Personendaten während ihrer Imagination, Hervorbringung, Bearbeitung und Verarbeitung. Das wird in den anschließenden Kapiteln zu zeigen sein.

»Was mit Daten möglich ist«

Stellen Sie sich vor, Sie bummeln durch die Stadt. Sie gehen an verschiedenen Geschäften vorbei und erhalten vor einem Kosmetikgeschäft eine Nachricht auf Ihr Smartphone. Die Nachricht wurde von der installierten Earlybird App versendet und weist Sie darauf hin, dass eine Haarspülung gerade exklusiv für Sie als Earlybird Member um 20 Prozent heruntergesetzt ist – just in jenem Geschäft, vor dem Sie gerade stehen. Die Haarspülung ist zudem das neue Produkt einer Marke, die Sie bereits kennen und mögen.

So ähnliche Geschichten erzählt Nik, der Geschäftsführer von Earlybird Digital. Nik ist ein Datenunternehmer: Sein Business besteht darin, andere Unternehmen auf ihre brachliegenden Datenbestände aufmerksam zu ma-

chen und sie zu beraten, wie sie Daten gewinnbringend nutzen könnten. Die beschriebene Geschichte animiert ein Projekt, an dem Nik und das Mutterunternehmen Earlybird gerade arbeiteten, als ich dort meine Feldforschung betrieb. Andere Unternehmen in der Schweiz hatten diese Vision mittlerweile umgesetzt – wenn auch nicht mit jener Art der Personalisierung, die Nik vorschwebt. Beispielsweise hatten die Schweizerischen Bundesbahnen (SBB) die App MyStation im Angebot, welche die Bewegung von Pendlerinnen per Bluetooth-Verbindung im Bahnhof trackte, um ihnen den schnellsten Weg zum Zug oder zum gewünschten Geschäft zu zeigen. Die App konnte per Push-Nachricht auf Rabattangebote wie zum Beispiel einen Cappuccino mit Gipfeli hinweisen.

Um eine Pendlerin auf einen Cappuccino mit Gipfeli hinzuweisen, musste sich die SBB nicht allzu weit aus dem Fenster lehnen. Mit einer solch generischen Empfehlung kann nicht viel schief gehen. Denn: Pendler mögen morgens Gipfeli und Kaffee. Auch wenn Nik manchmal selber solche einfachen Beispiele macht, zählt immer das richtige Angebot zur richtigen Zeit am richtigen Ort für die richtige Person. Einem Minderjährigen morgens um neun Uhr einen Rabatt für einen Krug Bier in der lokalen Kneipe vorzuschlagen, sei kaum angebracht. Sie wollen ihre Member zum richtigen Zeitpunkt auf ein individuell passendes Angebot aufmerksam machen. Dazu braucht Earlybird Daten.

An Daten über ihre Member kommen, Daten über ihre Member erzeugen oder Daten über ihre Member aus bestehenden Daten ableiten. Mit anderen Worten: möglichst viel über ihre Member zu wissen. Das ist die zentrale Mission des Digitalisierungsprojekts von Earlybird. Sie folgen damit dem von Marion Fourcade und Kieran Healy (2016) konstatierten »data imperative«, ein digitales, datengetriebenes Unternehmen zu werden. Wie Earlybird aber aus Erfahrung lernen musste, ist dieses Unterfangen aber nicht so einfach umzusetzen. Verschiedene Versuche der Datenerhebung und -auswertung sind auch schon gescheitert oder wurden auf Eis gelegt.

Nik sieht es als seine unternehmerische Aufgabe, Earlybird und anderen Unternehmen aufzuzeigen, »was eigentlich mit Daten alles möglich ist«. Seine Herangehensweise besteht darin, mit Beispielanalysen zu demonstrieren, was sich konkret mit Daten machen lässt, wie man an (gute) Daten kommt und was gute Daten sind. Gleichzeitig zielen seine Bemühungen auch darauf ab, in Unternehmen ein »data mindset« zu etablieren, d.h. den Mitarbeiten-

den vor Augen zu führen, dass sie bei allem, was sie tun, immer auch an Daten denken sollen.

Daten sind für Nik und für Earlybird immer auch ein Zukunftsversprechen, das während meines Forschungsaufenthalts viele ihrer Tätigkeiten anleitet, aber gleichzeitig immer auch zu entgleiten droht. Sie wollen herausfinden, wie sie zu einem datengetriebenen Unternehmen werden können, wie sie an Daten kommen und was sie damit machen können. Earlybirds Digitalisierungsprojekt und Niks unternehmerische Tätigkeit schließt an einen breiten Digitalisierungsdiskurs an, der Daten als eine natürliche Ressource und als probates Mittel zur Lösung bekannter und auch unbekannter, zukünftiger Probleme sieht. »By calling upon a future that is imminent but always just beyond reach, what technologies can currently do is not as important as what they might yet do in the future« (Elish & boyd, 2018, S. 13). Pointiert ausgedrückt: In Niks Formulierung steckt das Versprechen, dass mit Daten alles Mögliche möglich ist. Das ist im Verhältnis zu seinen kleinen, anwendungsbezogenen Beispielen und Geschichten die große Story der Daten. Gleichzeitig ist bei Earlybird Digital aber auch klar, dass – auch wenn sich mit Daten viele verschiedene Zwecke erreichen lassen –, sie doch überzeugend demonstrieren müssen, was sich konkret mit Daten machen lässt, um Aufträge zu erhalten.

Klaus Schwab, der Begründer des World Economic Forum (WEF), hielt im WEF-Report »Personal Data: The Emergence of a New Asset Class« von 2011 fest, dass Personendaten¹ in naher Zukunft zu einer neuen Vermögensklasse werden würden:

personal data is generating a new wave of opportunity for economic and societal value creation. (...) As some put it, personal data will be the new ›oil‹ – a valuable resource of the 21st century. It will emerge as a new asset class touching all aspects of society. (Schwab, 2011, S. 5)

Daten als das neue Öl, das neue Gold oder das neue Geld zu präsentieren, ist erst mal eine grandiose Behauptung. Sie unterstellt, dass Daten als universelles Mittel fungieren können, mit dem sich alle möglichen Zwecke und

¹ Wenn im Weiteren von »Daten« die Rede ist, meine ich in der Regel Personendaten (wenn nicht anders vermerkt).

Ziele erreichen lassen und das in andere Formen von Kapital konvertierbar ist.²

Von Daten als dem neuen Öl oder dem neuen Gold zu sprechen, ist nur eine der prominenten Varianten, mithilfe von Naturmetaphern fremde, neue oder womöglich bedrohlich wirkende Technologien zu verstehen, zu »domestizieren« und sie in bestehende Weltsichten zu integrieren (Lupton, 2014, 2016, Kapitel 5).³ Eine weitere metaphorische Umschreibung von Big Data bemüht Bilder von Wasser, Liquidität und »Flow« (Nelms, 2014): In Daten kann man schwimmen – oder ertrinken. Daten »fließen« durch »Ökosysteme« (Schwab, 2011) und füllen Flaschen oder ganze »data lakes«:

If you think of a datamart as a store of bottled water – cleansed and packaged and structured for easy consumption – the data lake is a large body of water in a more natural state. The contents of the data lake stream in from a source to fill the lake, and various users of the lake can come to examine, dive in, or take samples. (Marr, 2018)

Gegenwärtige Metaphern, die das Wesen und den Umgang mit (Big) Data umschreiben, versprechen Potenziale und beschwören Gefahren. Die Vorstellung eines Meers an Daten verdeutlicht sowohl die Gefahr als auch das große Potenzial, jene Gefahr durch erfolgreiche Navigation zu meistern und der Daten mächtig zu werden: Entweder ertrinken oder schwimmen und »insights« gewinnen (Deloitte, 2018).

»Spear fishing«

Die Spannung zwischen dem großen Potenzial von Personendaten und der Frage, was konkret damit möglich ist, kommt bei Earlybird verdichtet in der Metapher des »spear fishing« zum Ausdruck.⁴ Die »Datenleute« verfügen gemäß eigener Einschätzung nicht über das notwendige »Domänenwissen«⁵ und müssten deshalb Datenanalysen »ins Blaue hinein« machen. Deshalb

² Vgl. Deutschmann (2000) zu Geld als absolutem Mittel.

³ Die verbreitete Verwendung solcher Metaphern verweist auf die nicht abgeschlossenen Deutungsprozesse und die interpretative Flexibilität (Pinch & Bijker, 1984) der Big-Data-Technologie. Auch wenn Madeleine Elish und danah boyd (2018) konstatieren, wie das Schlagwort Big Data aufgrund einer Assoziation mit Überwachung und Privatsphäreverletzungen bereits anderen Begriffen wie Artificial Intelligence oder Machine Learning weichen muss.

sind sie auf diejenigen angewiesen, die über spezifisches Wissen über mögliche Problemlagen oder Datenanwendungsfälle verfügen, d.h. Wissen darüber, wonach sie in den Daten suchen sollen. Simon, der Senior Data Scientist bei Earlybird, und sein Team würden zwar über das technische Knowhow der Datenauswertung verfügen, aber das Wissen über die Domäne verortet er bei Earlybirds Marketingleuten. In einer Sitzung präsentiert er einige Kennzahlen aus der Trackingdatenbank der App. Er fügt dann aber an, dass »seine Jungs« nicht einfach Zahlen generieren sollten. Das Marketing müsse sagen, welche Zahlen sie für ihr »daily business« brauchen. Max, der strategische Geschäftsführer, schlägt vor, dass sie doch mal schauen könnten, was ihnen in den Daten »auffällt«, da das Marketing vielleicht gar nicht wisse, was interessant sein könnte. Und auch Nina, die für den Kontakt mit den Banken zuständig ist, hier aber als Vertreterin des Marketings spricht, möchte erst mal von Simon und seinem Team wissen, was es denn »alles gibt«. Simon erklärt, dass sie das schon machen könnten, dass das aber die »Luxusvariante« sei, die er als »spear fishing« bezeichnet: Dazu müsse man jemanden bezahlen, der explorativ die Daten auswerte und »einfach mal mit dem Speer ins Wasser sticht«. Max schlägt einen Experten für Datenanalysen vor, der Earlybird Anfang Oktober beraten hatte. Dieser verfüge über viel Erfahrung und könnte dadurch den »Teich« vielleicht etwas verkleinern.

Die vermeintlich nebenbei geäußerte Aussage »was mit den Daten möglich ist« erfasst die Komplexität der Problemlage nicht ganz. Es scheint alles oder zumindest vieles möglich zu sein, doch was soll damit konkret geschehen? Entgegen der großen Erzählung sind Daten nicht per se bedeutsam und wertvoll für Earlybird Digital: Um als Unternehmen erfolgreich zu sein, müssen sie zeigen, was mit ihnen *konkret* möglich ist.

-
- 4 Nicht zu verwechseln mit »spear phishing« als Praxis von Cyberkriminellen.
 - 5 Auf den Begriff der »Domäne« macht mich erstmals Dani aufmerksam. Er studiert bei Simon an der Hochschule für Informatik und ist bei Earlybird Digital angestellt, um ein Empfehlungssystem für Earlybird zu entwickeln. Er erklärt mir einige Probleme, die sich bei der Entwicklung von Recommendern stellen. Eines dieser zentralen Probleme ist die »Domäne« oder auf Englisch »domain«. Er erklärt mir, dass verschiedene Domänen unterschiedlich funktionieren würden und unterschiedliche Zielsetzungen und Businessmodelle hätten. Er verdeutlicht die Differenz am Beispiel von Netflix und Blendle, einer Webseite, die Newsartikel verschiedenster Printmedien für Nutzerinnen personalisiert auswählt. Beide Webseiten geben ihren Nutzerinnen Empfehlun-

Cornelius Puschmann und Jean Burgess (2014) identifizierten in Online-Newsbeiträgen zum Thema Big Data zwei Grundmetaphern: Große Datensmengen als natürliche Kraft (Wasser, Flut), die es zu kontrollieren gilt, und Big Data als zu konsumierende Ressource (Öl, Gold). In ihrem Artikel kritisieren sie die Unzulänglichkeit dieser Metaphern. Einerseits verdecke die Vorstellung von *Big Data als natürlich vorkommender Ressource* die datenerzeugenden Akteure. Daten werden nicht entdeckt, sondern geschaffen. Andererseits sei der Wert von Daten nicht für alle Akteure derselbe. Wert ist den Daten nicht inhärent. Er wird erst im Prozess der Analyse eingeschrieben:

Suggesting that the intrinsic meaning of data is, like nuggets of gold, already there, just waiting to be uncovered, means distancing the interpretation from the interpreter and her subjectivity. Data can be relied on as a driver of economic growth only if its value is both predictable and stable rather than the result of ongoing interpretation and negotiation, but the value of big data seems extremely difficult to predict. (Puschmann & Burgess, 2014, S. 1699)

Die Metapher von *Daten als zu konsumierender Ressource* unterschlage zudem die Interpretationsarbeit, die nötig ist, um Daten Sinn abzuringen. Was die Daten bedeuten, sei vermeintlich evident, »requiring no or very little conscious interpretation or reflection« (Puschmann & Burgess, 2014, S. 1700).

»Der Datenschatz«

Der zukünftige Erfolg oder Misserfolg von Earlybird Digital ist abhängig davon, ob sie Daten konkret nutzbar machen können. Dabei besteht die Annahme, dass Daten nicht nur den Erzeugern – die mit Wissen über den lokalen Kontext der Datenerzeugung ausgestattet sind –, sondern auch unbeteiligten Dritten wie Earlybird Digital Einblick in ihre Geheimnisse gewähren können. Earlybird Digital ist aber keineswegs naiv: Das Unternehmen weiß, dass es Daten erst angemessen verstehen kann, wenn es den Kontext ihrer Entstehung und die Bedürfnisse der Kundinnen versteht.

gen ab: Filme und Serien im Fall von Netflix; Nachrichtenartikel im Fall von Blendle. Die Domäne »Movie« unterscheidet sich aber von der Domäne »News« insbesondere dadurch, dass Newsartikel im Gegensatz zu Filmen schneller »verfallen« würden. Um einen guten Recommender zu bauen, braucht es Wissen über die Domäne, in welcher der Recommender eingesetzt werden soll.

Nik erzählt mir, dass vielen Firmen nicht bewusst sei, auf welchem »Datenschatz« sie säßen, was eine große Chance für Earlybird Digital sei. Gerade KMU ohne eigene Datenabteilungen nehmen ihre Daten-Dienstleistungen in Anspruch oder sind zumindest daran interessiert. Ein Beispiel für ein solches Unternehmen ist Hofstetter, ein lokales Kleidergeschäft mit fünf Filialen. Hofstetter wendet sich an ein gehobenes Publikum, das sich gerne in einem persönlichen Gespräch beraten lässt. Das Geschäft setzt auf wiederkehrende Kundinnen, die sich für eine Verkaufsberatung bei ihren bevorzugten Verkäuferinnen anmelden können. Earlybird Digital erhielt den Auftrag, explorativ aufzuzeigen, welche Daten überhaupt vorhanden seien und was damit möglich sei.

Die Metapher des Datenschatzes verweist auf die Arbeit, die notwendig ist, um den Schatz zu heben und zu sichten. Im Januar 2016 fahren ein Verkäufer von Hofstetter, Nik, Mike und ich als Beobachter nach Uri zum Hersteller des Kassensystems von Hofstetter. Fünf Stunden lang erklärt uns Stefan anhand eines komplexen Schemas die Datenstruktur: Kundenprofile, Lagerbestände, Warenzu- und -abflüsse, Einkaufs- und Verkaufspreise, Farben, Größen, Schnitte und wie alles zusammenhängt. Mike und Nik stellen Fragen zu den vielen verschiedenen Datenfeldern und Variablen. Stefan gibt mal mehr, mal weniger bereitwillig Antwort: Die Sitzung soll zwar einerseits Earlybird Digital helfen, die Datenstruktur und die Daten zu erschließen. Andererseits scheint Stefan aber auch abzuwägen, was er alles hergeben soll. Immer wieder kommt die Antwort, dass ein bestimmtes Feld für Earlybird Digital nicht relevant sei. Die Daten gehören zwar dem Kunden (d.h. Hofstetter), die Datenstruktur aber nicht: »Das geben wir eigentlich nicht raus«, meint Stefan. Den Datenschatz zu heben umfasst sowohl, die Daten und ihre Struktur zu verstehen, als auch deren Besitz- und Eigentumsverhältnisse zu verhandeln.

Die Metapher des Datenschatzes bezieht sich auf den Wert, der potenziell in den Daten steckt. Sie dient Nik als Marketingargument beziehungsweise als Versprechen an die Unternehmen, eine Ressource für sie nutzbar zu machen, die sie eigentlich bereits besitzen. Das Unternehmen selbst hat den Datenschatz angehäuft. Doch er blieb bisher unbemerkt und erschließt sich erst dem sachkundigen Blick der Experten.

Die Metapher des Datenschatzes suggeriert, dass Daten inhärent wertvoll seien. Im praktischen Verständnis von Earlybird Digital ist aber klar, dass der Wert der Daten nicht einfach so gegeben ist. Zuverlässig funktionierende

Infrastrukturen und Datenbanken, ein Blick für Daten sowie lokales Wissen sind notwendig, um den Schatz anzuhäufen, zu entdecken und schließlich als Ressource für weitere Anwendungen zu erschließen. Darin gerade besteht die Leistung von Earlybird Digital: den Unternehmen bei der In-Wert-Setzung von Daten behilflich zu sein.

Sowohl die Metapher der Datenflut als auch die Metapher von Daten als Resource geben vor, Daten seien unveränderlich, kontextunabhängig und hätten einen inhärenten Wert. Daten (beziehungsweise Big Data in der Untersuchung von Puschmann und Burgess) erscheinen als generalisiertes Mittel, mit dem sich beliebige Zwecke erreichen lassen. Solche Metaphern übertreiben, was mit Daten möglich sei. Wären Daten unabhängig vom Kontext ihrer Entstehung, könnten sie auch in ganz anderen Kontexten gewinnbringend sein – ihr inhärenter Wert bliebe stabil, auch wenn die Daten reisen.

Daten gelten als ein Mittel mit dem sich beliebige Zwecke erreichen lassen. Daten/Algorithmen – verstanden als sozio-technisches Ensemble (Gillespie, 2016) – erscheinen als »universelle Mittel«, »um komplizierte Probleme zu lösen und objektive vertrauenswürdige Entscheidungen herbeizuführen« (Häussling 2019, S. 332; siehe auch: Beer 2017). Der konkrete Nutzungswert der Daten ist aber oftmals zweitrangig (Sadowski, 2019, S. 4f.).⁶ Die Vision wertvoller Daten, kombiniert mit mächtigen Algorithmen, animiert ge-

6 Es ist keine neue Erkenntnis, dass sich der Nutzen von Daten erst in der Zukunft zeigen könnte. So heisst es beispielsweise in der verfassungsrechtlichen Überprüfung des deutschen Volkszählungsgesetzes von 1983: »Bei der Datenerhebung für statistische Zwecke kann eine enge und konkrete Zweckbindung der Daten nicht verlangt werden. Es gehört zum Wesen der Statistik, daß die Daten nach ihrer statistischen Aufbereitung für die verschiedensten, nicht von vornherein bestimmmbaren Aufgaben verwendet werden sollen; demgemäß besteht auch ein Bedürfnis nach Vorratspeicherung. Das Gebot einer konkreten Zweckumschreibung und das strikte Verbot der Sammlung personenbezogener Daten auf Vorrat kann nur für Datenerhebungen zu nicht-statistischen Zwecken gelten, nicht jedoch bei einer Volkszählung, die eine gesicherte Datenbasis für weitere statistische Untersuchungen ebenso wie für den politischen Planungsprozeß durch eine verlässliche Feststellung der Zahl und der Sozialstruktur der Bevölkerung vermitteln soll. Die Volkszählung muß Mehrzweckerhebung und -verarbeitung, also Datensammlung und -speicherung auf Vorrat sein, wenn der Staat den Entwicklungen der industriellen Gesellschaft nicht unvorbereitet begegnen soll. Auch wären Weitergabe- und Verwertungsverbote für statistisch aufbereitete Daten zweckwidrig.« (Bundesverfassungsgericht, 1983)

genwärtige und zukünftige Projekte der Datensammlung und der Digitalisierung (Elish & boyd, 2018).⁷ Marion Fourcade und Kieran Healy diagnostizieren einen »data imperative«, ein institutionalisierter, organisatorischer Mythos, welcher Organisationen motiviert, Daten zu sammeln, selbst wenn weitgehend unklar ist, ob und wie diese Daten genutzt werden könnten.

It does not matter that the amounts [of data] collected may vastly exceed a firm's imaginative reach or analytic grasp. The assumption is that it will eventually be useful, i.e. valuable. (Fourcade & Healy, 2016, S. 13)

Daten werden trotz fehlender konkreter Vorstellungen, Projekte oder Anwendungen erst mal »gesammelt« – in der vagen Voraussicht, dass sie sich später als nützlich erweisen werden.

Daten entstehen nicht in einem Vakuum, sondern in verschiedenen sozialen Kontexten (boyd & Crawford, 2012). Diese Kontextgebundenheit wird als eine De-Essenzialisierung von Daten verstanden. Kitchin (2014) erklärt – trotz einer Aufzählung verschiedener relativ kontextunabhängiger Typologien von Daten –, dass Daten immer gerahmt sind und diese Rahmen für eine soziologische Analyse unbedingt miteinbezogen werden müssen (siehe auch Borgman 2015, S. 18):

While many analysts may accept data at face value, and treat them as if they are neutral, objective, and pre-analytic in nature, data are in fact framed technically, economically, ethically, temporally, spatially and philosophically. Data do not exist independently of the ideas, instruments, practices, contexts and knowledges used to generate, process and analyse them. (Kitchin, 2014, S. 28)

Gegen die Sichtweise von »data before the fact« lancieren Lisa Gitelman und Virginia Jackson (2013) ihre viel zitierte Kritik an der scheinbaren Rohheit von Daten als Widerspruch in sich (Gitelman, 2013). Die Metapher der Rohdaten harmoniert mit der von Daniel Rosenberg (2013) beschriebenen rhetorischen Funktion von Daten und linearen Vorstellungen von Wissenserzeugung:

At first glance data are apparently before the fact: they are the starting point for what we know, who we are, and how we communicate. This shared sense

⁷ Nick Couldry und Joseph Turow (2014) weisen darauf hin, dass die epochale Rede von Big Data auch eine Zumutung sein kann. Proponenten von Big Data würden sogar Organisationen und Branchen wie Werbung und Marktforschung »belehren« wollen, in denen Big Data schon längst ein wesentlicher Bestandteil ist.

of starting with data often leads to an unnoticed assumption that data are transparent, that information is self-evident, the fundamental stuff of truth itself. If we're not careful, in other words, our zeal for more and more data can become a faith in their neutrality and autonomy, their objectivity. (Gitelman & Jackson, 2013, S. 2f.)

Auf der einen Seite scheinen Daten als rohe und universell anwendbare Resource (Anderson, 2008), einen objektiven »view from nowhere« (Haraway, 1988; Elish & boyd, 2018) zu ermöglichen. Andererseits erscheinen Daten niemals roh (Gitelman, 2013), sondern immer »gekocht« (Biruk, 2018; Bowker, 2013; Boellstorff, 2015), lokal (Loukissas, 2019) oder situiert (Elish & boyd, 2018).

Anstatt die Rede von der Rohheit der Daten als Übertreibung abzutun, zeigt sich im Feld, dass die damit verbundene Potenzialität der Daten, d.h. ihre potenzielle Formbarkeit zu vielen verschiedenen Anwendungsfällen wichtig und folgenreich ist. Mit Daten scheint alles Mögliche möglich zu sein. Andererseits erweist es sich für Earlybird und Earlybird Digital als praktisches Problem, etwas Konkretes mit Daten möglich zu machen. Wie deren Herangehensweise zeigt, sind auch sie – ähnlich wie Ethnografen – darauf angewiesen, sich Kontext- und Domänenwissen anzueignen, um die Daten und ihr Potenzial für Anwendungsfälle zu verstehen. Den Praktikern im Feld ist (zumindest im Fall von Earlybird) im konkreten Fall bewusst, dass die Daten erst einmal gebändigt werden müssen, damit sie ihren Dienst als Daten tun können.

Ethnografische Arbeiten zeigen, welche Reibungen während der Datenproduktion und -auswertung entstehen können: Dawn Nafus (2014) kritisiert prominente Datenmetaphern, um aufzuzeigen, dass Daten im Prozess ihrer Erzeugung und Auswertung um einiges widerspenstiger sind, als prominente Beschreibungen nahelegen. Sarah Pink et al. (2018) benutzen den Begriff der »broken data«, um diese Widerspenstigkeit und Eigensinnigkeit von Daten (in ihrem Beispiel Daten der Selbstquantifizierung) zu benennen. Sie wollen mit ihrer konzeptuellen Metapher der »broken data« die Aufmerksamkeit darauf lenken, dass Daten und ihre Infrastrukturen in einem permanenten Prozess der Beschädigung (»data as ongoingly being broken« (2018, S. 10)) stehen. Es geht ihnen aber nicht so sehr darum, den Datenbegriff zu demystifizieren, sondern vor allem den Fokus auf die Reparatur und den Unterhalt sowie die Kontingenzen von Daten und Infrastrukturen zu lenken (siehe z.B. Ensmenger (2014) für den Fall von Software im Allgemeinen). Anissa Tan-

weer et al. (2016) beschreiben die Arbeit mit Daten als einen Prozess von »breakdown und repair«. Diese ethnografischen, die Datenrhetorik kritisierenden, Arbeiten verdeutlichen, dass Daten im Alltag von Praktikerinnen um einiges widerspenstiger sein können. Sie verweisen auf die Hinterbühne der Datenpraktiken und damit auch auf das bei Earlybird und Earlybird Digital identifizierte Problem, aus den vielversprechenden Daten überhaupt etwas machen zu können. Das Potenzial von Daten ist im Alltag von der Schwierigkeit der tatsächlichen Anwendung beziehungsweise dem Widerstand der Daten geprägt, sich produktiv zu zeigen. Die Metapher der rohen Daten ist aus methodologischer Hinsicht sicherlich zu kritisieren. Für die Anwendrinnen ist es aber vor allem ein praktisches Problem, aus Daten Rohdaten zu machen, die sich für weitere und womöglich unvorhergesehene Zwecke produktiv machen lassen.

Trotz der Erkenntnis der radikalen Situiertheit und Widerspenstigkeit von Daten ist nicht von der Hand zu weisen, dass digitale Personendaten in der Praxis oftmals tatsächlich als »immutable mobiles« funktionieren und ihre Situiertheit transzendieren. Bruno Latour weist darauf hin, dass Daten als Objekte fungieren, »which have the properties of being mobile but also immutable, presentable, readable and combinable with one another« (2011, S. 26).⁸ In Bezug auf digitale Daten machen diese Eigenschaften gerade ihre Datenhaftigkeit aus: Oftmals lassen sie sich von ihren lokalen Kontexten der Erzeugung ablösen und mit anderen Datensätzen kombinieren, so dass sie zumindest potenziell auch in ganz anderen Kontexten angewendet werden können. Der soziologischen/ethnografischen Ablehnung einer übersteigerten Rhetorik von Rohdaten steht gegenüber, dass genau darin deren Stärke besteht: Sie können von lokalen Kontexten abstrahieren. Daten *sind* in der

8 Siehe auch Madeleine Akrich (1992, S. 425f.): »Wenn technische Objekte einmal stabilisiert sind, werden sie Instrumente des Wissens. Wenn also eine Elektrizitätsgesellschaft unterschiedliche Tarife für viel und wenig verbrauchende private Benutzer, für Werkstätten und für Industrieverbraucher ansetzt, findet sie Wege, verschiedene soziale Schichten zu charakterisieren und zu identifizieren. Wenn sie auch Kategorien wählt, die in anderen sozioökonomisch-politischen Netzwerken verwendet werden, kann das von ihr produzierte Wissen ‚exportiert‘ werden. ›Daten‹ können also aus dem Netzwerk gezogen und an einen anderen Ort übermittelt werden, z.B. zu Ökonomen, die an einer Beziehung zwischen den Energiekosten oder dem Bruttonsozialprodukt und dem Konsum interessiert sind. Die Konversion soziotechnischer in reine und einfache Fakten hängt jedoch von der Fähigkeit ab, technische Objekte in Black Boxes umzuwandeln.«

Praxis mobil: Sie werden an anderen Orten, zu anderen Zeiten und von anderen Akteuren neuen Zwecken zugeführt. Das scheint Daten überhaupt zu Daten zu machen. Wie auch Zahlen (Heintz, 2010; Porter, 2001) scheinen Daten einfacher als Sprache zwischen verschiedenen lokalen Kontexten reisen zu können.

Einen entscheidenden Punkt von Gitelman/Jackson (2013) sehe ich in einer Formulierung in der Einleitung zu ihrem Band:

the seemingly indispensable misperception that data are ever raw seems to be one way in which data are forever contextualized—that is, framed—according to a mythology of their own supposed decontextualization. (Gitelman & Jackson, 2013, S. 5f.)

Das Rätsel der Daten liegt in dieser Kontextualisierung als dekontextualisiert. Die Datenkritik verweist auf diesen Widerspruch: Daten sind immer gleichzeitig zu wenig und zu viel. Zu wenig, weil Daten (genauso wie auch Zahlen) die Kontinuitäten des Alltags auf diskrete Einheiten reduzieren und alles, was nicht in dieses Schema passt, ignorieren und unsichtbar machen. Daten sind aber auch immer zu viel, weil sie vor überschüssigen Sinngehalten triefen. Sie verweisen immer auch auf individuelle, soziale, technische, ökonomische und historische Kontexte ihrer Erzeugung. Gerade weil diese Kontexte der Datenerzeugung verschleiert werden, können Daten überhaupt als Daten funktionieren und zu anderen, neuen und nicht ursprünglich vorgesehenen Zwecken eingesetzt werden.

Daten – insbesondere Personendaten – verweisen immer auf »mehr«: Auf soziale Kontexte, in denen sie entstanden sind, auf vergangene und zukünftige Verhaltensweisen von Individuen oder auf Geheimnisse und Erkenntnisse, die in den Daten stecken. Gerade in diesen Sinnüberschüssen besteht ihre Faszination, die sich der Datenunternehmer Nik zu Nutze macht. Gleichzeitig wird dieser Sinnüberschuss aber auch von seinen Data Scientists weggearbeitet. Die Ambivalenz von Daten erweist sich praktisch gesehen als produktiver Widerspruch, an den die unternehmerischen Bemühungen der Erzeugung und In-Wert-Setzung von Personendaten anschließen können.