

Alle Rechte beim Urheber.
Abdruck nur gegen Belegexemplar.

Datenkrake Wissenschaft?

Daten sind im Begriff, zu einem grundlegenden Substrat der Wissenschaft zu werden und eine legitimatorische Position einzunehmen, wie sie bisher nur Fakten eingeräumt wurde. Sie schicken sich an, im Erkenntnisprozess verbindlicher noch gegeben zu sein als Sinnesdaten. Daten, nicht Tatsachen sind inzwischen gängige Münze der Wissenschaft.

Viele Wissenschaftler können der korrelativen Versuchung nicht widerstehen und lassen ungeprüfte Daten durch die mächtigen Mühlen der Statistikprogramme laufen. Big Data scheint die Versprechungen der Moderne einzulösen, eine Herrschaftsform auf unbestechlicher und anti-autoritärer Grundlage zu errichten. Im wertfreien Datenuniversum, das quantitativ in Bits und Bytes messbar ist, umgibt die Daten eine Aura sozialer Neutralität. Die Wissenschaft scheint in ihrer Bestimmung angekommen.

Doch die ‚Rohdaten‘ eines Datensatzes sind bei weitem nicht so roh und unberührt, wie es in ihrer Bezeichnung anklingt. Sie sind sehr wohl das Resultat einer Sammel- und Erzeugungsmethode, die zu bestimmten Zwecken ausgewählt wurde. Social Media-Plattformen wie Facebook, die das Verhalten ihrer Nutzer im Netz registrieren, haben einen entscheidenden Einfluss auf Art und Zusammensetzung der von ihnen gesammelten Daten. Indem Facebook nur ein begrenztes Repertoire an Verhaltensweisen zulässt, von denen das Unternehmen zudem ausgewählte begünstigt, dadurch dass es sie prominent platziert oder leicht umsetzbar macht, werden einige Daten besonders häufig generiert, während andere von vornherein ausgeschlossen sind. In diesem Sinne sind die implementierten Algorithmen selektiv und die von ihnen gesammelten Daten gemacht: Die Daten sind selbst eine Tatsache, die es zu untersuchen lohnt.

Auch darum werden Wissenschaftler zu Advokaten ‚freier Daten‘. Sie fordern freien Zugang zu jeglichen Daten im expandierenden Datenuniversum. Vor allem aber wollen sie verhindern, dass die Nutzerdaten in den Händen einiger weniger Unternehmen bleiben, deren Analysen dann öffentlich nicht überprüft werden können. Doch die Wissenschaftler fordern nicht nur die Freigabe von Daten, sie fordern die Freigabe aller Daten. Eine

Anonymisierung kommt für sie nicht in Frage. Vielmehr reklamieren sie für sich das öffentliche Vertrauen, mit den Daten verantwortungsvoll umzugehen.

Das hat gute Gründe. Erstens lässt sich eine Anonymisierung kaum gewährleisten. Sie ist immer relativ zu Identifizierungstechnologien. Anhand so genannter Quasi-Identifikatoren wie dem Geschlecht, dem Geburtsdatum und der Postleitzahl lassen sich bereits 87 Prozent der Bevölkerung identifizieren. Je mehr Datensätze es gibt, umso sicherer kann man schließen, wer was klickt, tut, beabsichtigt. Zweitens verzerren schon die leisesten Anonymisierungsversuche eines Datensatzes dessen Analyse. Die Mittelwerte verschieben sich, die Streuung und mithin die Korrelationskoeffizienten verändern ihre Größe infolge einer Anonymisierung.

Eines lässt sich mit großer Sicherheit zu Daten sagen: Einmal kreiert, sind sie für immer in der Welt. Daher werden sie immer mehr. Digitale Datenspeicher haben ein gigantisches Fassungsvermögen. Auf vier Zettabytes, das sind $4 \cdot 10^{21}$ Bytes, schätzte man 2013 das weltweit gespeicherte Datenvolumen. Dabei nehmen die Datenjäger und -sammler gerade erst so richtig Fahrt auf. Alles wird aufgezeichnet und gespeichert. Inzwischen werden mehr Sensoren produziert als Kartoffeln angebaut. Ganz vorne mit dabei: die Wissenschaft. Der Large Hadron Collider vom CERN steuert täglich 100 Terabytes bei, das sind 10^{14} Bytes; das Radioteleskop SKA soll ab 2024 am Tag mehr Daten generieren, als im Internet ausgetauscht werden, das sind immerhin mehr als 90 Exabytes oder $9 \cdot 10^{19}$ Bytes; die 1,5 Gigabytes ($1,5 \cdot 10^9$ Bytes) des menschlichen Erbgutes sind in eineinhalb Stunden sequenziert.

Die Genomik verspricht sich von der schleusenlosen Datenflut erhebliche Fortschritte. So ist beispielsweise eine Genvariante für Schizophrenie nicht nachweisbar in 3500 Genomen, sie wird aber signifikant in 35000 Genomen. Fragt sich nur, wie man an das Erbgut von 35000 Schizophrenen herankommt? Schließlich reagieren weite Teile der Bevölkerung äußerst sensibel, was zumindest die latenten Schattenseiten ihrer Basen-Triplets angeht. Dabei muss man gar nicht erst auf molekularbiologische Daten zurückgreifen, um sich ein so detailliertes wie umfassendes Bild von einem Menschen zu machen. Es sind bereits jetzt schon so viele personenbezogene Daten digitalisiert, dass dem Menschen der Rückzug ins Private verbaut ist. Auch ohne dass er seine Gedanken oder Gene preisgibt, lässt sich algorithmisch mit großer Treffergenauigkeit auf seine Gedanken - und irgendwann auch seine Gene - schließen.

inspective.

Das hat zur Folge, dass Dritte einen Menschen nicht nur besser kennen als er sich selbst. Auch das Menschenbild vollzieht einen grundlegenden Wandel: Der Mensch mutiert vom Subjekt zum Objekt, vom autonomen (Grund-) Rechtsträger, dessen künftige Handlungen einem freien Willen unterworfen sind, zu einem planbaren Datenträger, der die Handlungen abspielt, die sich auf ihm speichern lassen. Und die Handlungen beschränken sich nicht nur auf das Konsumverhalten. Es muss nicht gelingen, es reicht, dass es versucht wird!

In einer Datenwolke aus Adressen, Mitgliedschaften, Kontoverbindungen, Telefonnummern und Gesprächsdauern, aus Bildungsabschlüssen, Reisen, gekauften und ersteigerten Produkten oder aus IPs und aufgerufenen Webseiten lassen sich Profile von individuellen Menschen verdichten; und jedes neue Datum eröffnet den Profilen eine Dimension, in die eine Person zu ihrer Charakterisierung hineinprojiziert werden kann. Dieser Mosaik-Effekt, demzufolge Personen anhand unpersönlicher Daten identifiziert werden können, wird kontrapunktiert vom Effekt selbst-induzierter Verzerrungen, wonach die Auswertung von Big Data bei hinreichender Aufmerksamkeit der Netzgemeinde Aktivitäten auslöst, die nicht im Zusammenhang mit dem Thema, wohl aber im Zusammenhang mit der Thematisierung stehen. So ist es zu erklären, dass Google bei der epidemiologischen Beschreibung der Grippe weltweit teilweise völlig daneben lag. Noch ist die Auswertung von Big Data nicht so weit, dass sie etablierte Methoden und Theorien ersetzen könnte. Aber sie ist für Viele ein verheißungsvolles Versprechen.

Das ruft die Datarazzi der sozialen Netzwerke auf den Plan, die schon mit der Illusion, das Bedürfnis nach sozialer Zugehörigkeit und Verbundenheit zu befriedigen, eine Unzahl von Followern ködern und so zu einer Art psychologischer Kolonialisierung der Privatsphäre ansetzen. Die ist jetzt schon unumkehrbar. Freizeit und Arbeitswelt, ja der öffentliche Raum sind in einem Grade durchdigitalisiert, dass man nirgendwo mehr Mitglied werden oder einkaufen kann, ohne eine Spur zu hinterlassen in einer Datenbank, deren Inhalt verknüpft ist mit anderen Datenbanken und -quellen. Jede Internet-Aktivität füttert einen individuellen Avatar, dessen Identität der des Nutzers vorausläuft und leicht zu einer schweren Hypothek werden kann, wenn den algorithmischen Komponenten des Avatars mehr Glauben geschenkt wird als der aufrichtigen Selbstauskunft seines - unfreiwilligen - Erzeugers. Schließlich lügen Daten nicht. Das Subjekt ist mit seiner begrenzten Rationalität chancenlos gegenüber der unbegrenzten Rationalität algorithmischer Datenverknüpfungen.

inspective.

Es verschwindet hinter dem Projekt des Datensammelns und -verknüpfens.

Problematisch daran ist, dass das einzelne Subjekt ‚sein‘ Projekt nicht kontrolliert. Und es lässt sich nicht schützen vor Geisterfahrern auf der Datenautobahn. Alleine die im Netz verstreuten Daten machen Algorithmen zu einem mächtigen Instrument, einem Nutzer in Echtzeit Service- oder Produktangebote zu unterbreiten, noch bevor der sich seines Bedürfnisses danach bewusst wird. Genauso können Parteien gezielt politische Botschaften absetzen, weil das Thema demnächst für den Nutzer interessant werden wird. Die Politik setzt bei ihren Aktivitäten dann zusehends auf eine zielgenaue Analyse der Wählerschaft und wohltemperierte Botschaften statt auf rationale Überzeugungskraft.

Das delphische *Cognosce te ipse* haben Algorithmen übernommen; Luthers *Servum arbitrium* diejenigen, die Algorithmen programmieren oder kaufen können. Wer einem Marathonläufer just an der Stelle ein Sofa und gekühlte Getränke hinstellt, an dem sich bei ihm der innere Schweinehund meldet, mag das als Dienstleistung verkaufen, er stellt tatsächlich aber den noch bestehenden Gesellschaftsvertrag infrage, der auf dem Gedanken der Privatautonomie aufbaut. Denn er maximiert die Wahrscheinlichkeit, den Läufer von seinem ursprünglichen Vorhaben abzubringen. Der freie ungezähmte Wille fügt sich willenslos in vorgefertigte Bahnen. Unwillkürlich entbrennt dann ein Wettstreit, schneller und erfolgreicher das Verhalten und die Identität von Menschen zu manipulieren, seine Existenz einem übergeordneten - nicht mehr frei wählbaren - Zweck unterzuordnen. Das Subjekt ist dann endgültig Projekt geworden.

Das Subjekt-Projekt kann seine Privatsphäre gar nicht schützen - sie ist internet-öffentlich und kann von jedem eingesehen (und ausgewertet) werden, der Zugang zu den Daten hat. Je nach Profil sind Preise und Gebühren gestaffelt, werden Kredite limitiert oder Policen verweigert. Dem kranken alten Deutschen zeigt das Internet ein anderes Gesicht als der gesunden jungen Iranerin. Die Diskriminierung folgt der Profilierung wie ein Schatten. Das geht jetzt schon hin bis zu smarten Filter Bubbles, die verhindern, dass bestimmte Personen an Informationen gelangen können, die beispielsweise das herrschende System in Frage stellen könnten. Die Totalüberwachung nimmt sich da aus wie eine bloße Vorstufe der gegenwärtigen Per- und Konversionen der Gesellschaft.

Unternehmen und Behörden wissen heute schon mehr um eine Person, ihre Herkunft und Zukunft, als die Person selbst; und

sie wissen zusehends, wie sie dieses Wissen nutzen können, um die Person zu beeinflussen – nicht nur hinsichtlich des Einkaufens, sondern auch in Bezug auf ihre politische Einstellung, ihren Selbstwert und ihr Selbstverständnis und damit auf ihre gesellschaftliche Partizipation. Diese wachsende Machtfülle geht einher mit einer auseinanderklaffenden Asymmetrie der Macht zwischen denen, die die Daten haben, die sie auswerten können und denen, die sie freiwillig oder unfreiwillig liefern.

Angesichts einer so tiefgreifenden Transformation bleiben auch die Grundwerte und -rechte nicht unberührt. Die Freiheiten, auch die Wissenschaftsfreiheit, oder das Diskriminierungsverbot und ihre Durchsetzung sowie soziale Gerechtigkeit werden von Big Data herausgefordert und müssen neu gesichert werden. Vor allem müssen Wege in eine Gemeinschaft gefunden werden, in der ein Recht auf informationelle Selbstbestimmung unhaltbar geworden ist. Stattdessen braucht es eine diskursive Verortung dessen, was die Identität eines Menschen ausmacht und was uns ihr Schutz wert ist. Sonst dominiert weiter allein der Effizienzgedanke die datengetriebene Transformation, sei es in der Industrie, dem Energieverbrauch, dem Gesundheitswesen oder der inneren Sicherheit. Dann heißt es die Leistung verbessern, Kosten sparen und weiter wachsen. Wozu? Um noch mehr kaufen zu können. Fragt sich nur was, gibt es doch jetzt schon mehr Mobiltelefone als Menschen auf der Welt. Vielleicht etwas, womit man noch mehr und noch detailliertere Daten sammeln könnte?

Von der Wissenschaft darf man mehr erwarten als Daten sammeln. Die Leviathan-Bezwingerin sollte es auch mit Big Data aufnehmen können. Noch sollten sich unter ihrem Schild genug renitente Selbstdenker finden. Welch historischer Irrtum, wenn die sich als bloße Datenkrake entpuppten!