

Bachelorarbeit

Soziale Daten in der EU

von Maria Chlastak

betreut von Dr. Karl Ucakar

im Fach Politikwissenschaft

Sommersemester 2020

Abstract:

Soziale Daten können als epistemologische Grundlage jeder Metatheorie verstanden werden die das Soziale berührt. Trotz empirischer Existenz war diese Datenart bisher unbeachtet. Ich zeige das Vorkommen in wissenschaftlichen Arbeiten und verorte soziale Daten im Kontext der Sozialwissenschaften, der Digitalisierung und der Prüfung nach Foucault. *Soziale Daten sind erfassbar, wiederverwendbar, historisch bedingt und eine Summe aus personenbezogenen Daten, in dem Sinn, dass soziale Daten in ihrer Entstehung und durch ihren Inhalt auf einzelne Personen zurückführbar sind.* Grenzfälle sozialer Daten entstehen bei der Datenverarbeitung, bei der Feststellung von Fakten und bei privaten Daten. Zahlreiche Beispiele konstituieren soziale Daten in der EU. Die Datenschutz-Grundverordnung, die europäische Datenstrategie und die Infrastruktur Zenodo werden anhand der Theorie analysiert. Der Text ist kritisierbar und kann als Theorie für weitere Forschung mit sozialen Daten verstanden werden.

0.		
Einleitung		3
1.		
Was sind soziale Daten?		4
1.1.	Vorkommen	5
1.2.	Kontext	6
	1.2.1. Sozialwissenschaften	7
	1.2.2. Digitalisierung	8
	1.2.3. Die Prüfung	9
1.3.	Definition	10
	1.3.1.	
	Charakteristiken	11
	1.3.1.1. sozial	11
	1.3.1.2. erfassbar	11
	1.3.1.3. wiederverwendbar	12
	1.3.1.4. historisch	12
	1.3.1.5.	
	Summe	13
	1.3.2. Anwendungsfall "Big Social Data"	14
1.4.	Beispiele sozialer Daten	15
1.5.	Grenzfälle	16
	1.5.1. Datenverarbeitung	17
	1.5.2. Fakten	17
	1.5.3. Private Daten	18
2.		
Soziale Daten in der EU		20
2.1.	Die Datenschutz-Grundverordnung	20
2.2.	Europäische Datenstrategie	22
2.3.	Zenodo	25
2.4.	Zusammenfassung	27
Literaturverzeichnis		28

0.

Einleitung

Spätestens seit der Datenschutzgrundverordnung ist die europäische Bevölkerung zunehmend mit der Bedeutung von sozialen Daten konfrontiert. Das Internet hat sich durch Cookie-Banner spürbar verändert. Neue Jobs und Verantwortliche wurden geschaffen. Rechte garantieren uns vollen Zugriff auf unsere persönlichen Daten, welche Online-Unternehmen für uns verarbeiten. Forscher_innen erheben Daten in Größendimensionen, die im Analog-Zeitalter undenkbar waren. Dadurch wird sichtbar, was sonst unerfasst geblieben wäre. Das bietet ungeheure Potenziale für die Wissenschaft, insbesondere für die Sozialwissenschaften.

Datenmanagement, Datenqualität und die Präkonditionen maschineller Ergebnisse sind ständige Begleiter der Data Science, die sich durch alle wissenschaftlichen Disziplinen zieht. Je nach Fachgebiet sind andere Daten relevant. Dazu sammeln Verwaltung und die Exekutive Daten in ihrem Hoheitsgebiet. Firmen generieren und verarbeiten Daten mit dem Ziel der Wertschöpfung. Menschen behalten Daten, wie Fotos, Videos oder Nachrichten als Erinnerung. Dateiformate sind so vielfältig wie die Art der Daten selbst. Immer wieder werden neue Formate erfunden, um spezielle Inhalte angemessen aufzuzeichnen. Noch nie war die Mathematik so nah an den Sozialwissenschaften und dennoch lässt dieser Transformationsprozess eher diverse neue Internet-Disziplinen entstehen, anstatt die bestehenden einzubinden.

Die Differenzierung von Daten ist genauso notwendig, wie eine gemeinsame Debatte, was Daten für das gesellschaftliche Leben bedeuten. Ähnlich wie die neuen Daten Chancen schaffen, laufen sie Gefahr, einseitig von einzelnen Akteur_innen ausgenutzt zu werden. GAFA (Google, Apple, Facebook, Amazon) sind jene Oligopole, die das Internet der westlichen Hemisphäre bestimmen. Sie sind Konzerne, die sich bei der Verbreitung des Internets durchgesetzt haben. Dieselben beeinflussen politische Akteur_innen weltweit. Ich schreibe diese Arbeit als Europäerin und nutze offizielle europäische Quellen um soziale Daten zu konstituieren. Das scheint notwendig, weil diese Daten im Forschungsprozess

verarbeitet werden, ohne dass diese ausreichend besprochen wären (Anzinger, Hamacher & Katzenbeisser, 2013; Atzmueller, 2012; DeMaris, 2004; Gross, 2011; Guo, Vargo, Pan, Ding & Ishwar, 2016; Karimi, Tulloch, Brown & Hockings, 2017). Soziale Daten sind zentral für die Digitalisierung, aber die Generierung von Wissen ist kein neutraler Prozess: Es ist anzunehmen, *“daß es keine Machtbeziehung gibt, ohne daß sich ein entsprechendes Wissensfeld konstituiert, und kein Wissen, das nicht gleichzeitig Machtbeziehungen voraussetzt und konstituiert.”* (Foucault, 1977:39) In diesem Kontext sollen auch die europäischen Beispiele begriffen werden, die soziale Daten in Europa darstellen. Ich ziehe meine Theorie aus Gesprächen, Dokumenten und Analysen zum Thema. Durch den Begriff *soziale Daten* hoffe ich das Datenchaos ein Stück weit zu entwirrt, damit konkretere Policies und Gedanken formuliert werden können. Dieses Dokument ist kritisierbar und wird je nach Interesse weiter überarbeitet.

Im ersten Teil der Arbeit werde ich soziale Daten genauer erfassen, definieren und Beispiele nennen. Es soll ein klares Bild darüber entstehen, was soziale Daten sind und wo ihre theoretischen Grenzen liegen. Dabei versuche ich verständlich zu schreiben, um die Zugänglichkeit zu erhöhen. Im zweiten Teil wende ich die Theorie auf drei ausgewählte europäische Beispiele an: Erstens, die Datenschutzgrundverordnung, welche als regulatorische Grundlage für Privatsphäre gesehen werden kann. Zweitens, die europäische Datenstrategie, die kommende Aktionen der Europäischen Union bestimmen wird. Drittens, die Infrastruktur Zenodo, die als freies, wissenschaftlich genutztes Data-Center vom CERN betrieben wird. Die Beispiele sind im Zugang maximal unterschiedlich und ergeben doch ein gemeinsames europäisches Bild sozialer Daten. So biete ich eine mögliche Antwort auf die Forschungsfrage: Wie konstituieren sich soziale Daten in der EU?

1.

Was sind soziale Daten?

Um diese Frage zu beantworten, recherchiere ich im Angebot der Universitätsbibliothek Wien, indem ich es nach Relevanz durchsuche. Die Suchbegriffe "Soziale Daten" und "social data", inklusive Abwandlungen, führten mich zu keiner akademischen Diskussion, die mir

theoretische Ansätze liefern würde, welche soziale Daten näher definieren. Nachdem ich jeweils über 100 Werke nach Definitionen gescannt habe, beginne ich, nach überschaubarem Erfolg, einzelne Disziplinen und Themen herauszufiltern, die sich mit Technik, Humanwissenschaften oder Sozialwissenschaften befassen. Ich wollte wissen, in welchen Bereichen soziale Daten Erwähnung finden. Dabei beachte ich peer-reviewed Journals und Dokumente von öffentlichen Stellen. Soziale Daten (social data) werden in beiden Sprachen wissenschaftlich gebraucht, ohne dass diese näher bestimmt werden.

1.1.

Vorkommen

Soziale Daten sind deutlich stärker im englischsprachigen Raum verbreitet. "Social data" können im Rahmen der Digitalisierung erfasst werden. Es geht um statistisch verwendbare Daten, die vor allem in den Sozialwissenschaften, der Humanmedizin, im Kontext von Citizen Science und in technischen Disziplinen auftreten. Im deutschen Sprachraum findet man hauptsächlich Arbeiten aus der Medizin, in denen soziale Daten zur Anwendung gebracht werden. Selbst in den Sozialwissenschaften stehen soziale Daten nicht im Mittelpunkt der Arbeit, dennoch werden sie zur Generierung von Forschungsergebnissen gebraucht.

DeMaris erklärt in seinem Buch *"Regression with Social Data: Modeling Continuous and Limited Response Variables"*, wie man soziale Daten statistisch anwendet. Die Arbeit richtet sich an Student_innen der Sozial- und Verhaltenswissenschaften und an Daten Analyst_innen, die mit sozialen Daten arbeiten. (DeMaris, 2004:xvi.) Soziale Daten werden statistisch angewandt und nicht theoretisch erfasst.

Fünf Jahre später veröffentlicht Mitchell einen kurzen Essay im Gebiet Computer Science: *"Mining Our Reality"*. Die Arbeit habe ich im Literaturverzeichnis eines Essays von Atzmueller (Atzmueller, 2012) entdeckt. Die Konklusion von Mitchell ist bis heute relevant und erwähnenswert: *"Perhaps even more important than technical approaches will be a public discussion about how to rewrite the rules of data collection, ownership, and privacy to deal with this sea change in how much of our lives can be observed, and by whom. Until*

these issues are resolved, they are likely to be the limiting factor in realizing the potential of these new data to advance our scientific understanding of society and human behavior, and to improve our daily lives." (Mitchell, 2009:1645)

Weder Mitchell noch Atzmueller erklären das Phänomen sozialer Daten. Atzmueller versucht sich hingegen an einer anderen Definition: "Intuitiv verstehen wir unter Social Media daher internetbasierte Systeme und Dienste im ubiquitären Web, die als Medium soziale Daten menschlicher Interaktion und Kommunikation bereitstellen." (Atzmueller, 2012:132) Die Informatiker arbeiten mit sozialen Daten und sehen den Wissenszuwachs als Potenzial und Rechtsunsicherheit als Risiko.

Ein Sammelband des Springer-Verlags mit dem Titel "Schutz genetischer, medizinischer und sozialer Daten als multidisziplinäre Aufgabe" erscheint im Jahr 2013. Ein ganzer Teil des Buches widmet sich sozialen Daten. Gottschlich und Hinz erklären beispielsweise den "Wert von sozialen Strukturdaten aus ökonomischer Sicht" (Anzinger et al., 2013:87ff). Soziale Daten werden als Eigentum kontextualisiert. Ein Phänomen, das auch in sozialwissenschaftlichen Arbeiten auftritt. Klaus Pommerening schreibt im selben Buch einen Beitrag über "Personalisierte Medizin: Herausforderungen für den Datenschutz und die IT-Sicherheit" (Anzinger et al., 2013:21ff). Er sieht, dass Endgeräte nicht vertrauenswürdig sind, während immer mehr Daten anfallen, die durch die ärztliche Schweigepflicht geschützt sind. An adäquaten technischen Maßnahmen fehle es. (ebd.) Eine sozialwissenschaftliche Definition sozialer Daten ist im Sammelband nicht zu finden.

1.2.

Kontext

Sofern man eine akademische Debatte über soziale Daten erkennen kann, sieht man sie fachspezifisch und zumeist im technischen Métier. Sowohl in der Informatik, als auch in den Humanwissenschaften herrschen objektivistische Erkenntnisstrategien vor. Demgegenüber bieten sich Sozialwissenschaftler_innen auch konstruktivistische Strategien zur Analyse an. Ich zeige drei verschiedene Kontexte auf, die nicht getrennt voneinander denkbar sind: Erstens, die Sozialwissenschaften, in denen sich auch die Politikwissenschaft wiederfindet,

welche soziale Daten erforschen. Zweitens, die Digitalisierung, welche die Generierung von sozialen Daten fördert. Drittens, setzte ich soziale Daten in den Kontext der Prüfung nach Michel Foucault (1926-1984). Der Theoretiker beschäftigte sich unter anderem mit Macht und totalen Institutionen. Anhand der Prüfung wird sichtbar, wie soziale Daten einseitig zur Beschränkung des Menschen genutzt werden.

1.2.1.

Sozialwissenschaften

Mathematik und Philosophie wurden in der Antike gemeinsam betrieben. Dokumentation, Statistik und Mathematik werden erst in der neueren Geschichte wieder relevant, in welcher auch ein langsames Auftauchen der Sozialwissenschaften erkennbar ist. Obwohl die Erhebung und Bewertung von sozialen Daten schon länger in verschiedenen staatsnahen Institutionen stattfindet, sind die Daten meistens nicht für die Wissenschaft zugänglich. Sozialwissenschaftliche Ansätze tauchen in humanwissenschaftlichen und philosophischen Texten auf, wobei man sich selten auf große Datensätze berufen kann. Die Erhebung sozialer Daten ist aufwendig und kostspielig. Erst nach dem Zweiten Weltkrieg erleben die Sozialwissenschaften ihre Blütezeit in Europa. Die Differenzierung des Sozialen von den Naturwissenschaften scheint gelungen und es gibt Raum für perspektivische Forschung. Sozialwissenschaften werden an Universitäten gelehrt. Immer mehr Menschen interessieren sich für den Bereich. Es gibt immer mehr Studierende. Viele Wissenschaftler_innen erlauben größere Meinungs- und Methodenvielfalt, die wiederum tiefere Einblicke in die Natur des Sozialen geben. Die Ausdifferenzierung in verschiedene Disziplinen, die wiederum in verschiedene Teildisziplinen unterteilt ist, scheint funktional, doch bewirkt sie auch einen Informationsüberfluss, der durch das Internet verstärkt wird. Das führt zu Problemen bei der grundlegenden Vermittlung des Wissens an fachfremde Personen. Ob die Sozialwissenschaften mit ihren Disziplinen auch nach der Digitalisierung eine Rolle spielen, kann nur die Zeit zeigen, die aktiv von Wissenschaftler_innen gestaltet wird.

1.2.2.

Digitalisierung

In den 2000ern beginnt das Rennen um die Digitalisierung, welches die Welt in nur 20 Jahren massiv verändert hat. Für die einen ist Digitalisierung Prestige und Geschäft, für die anderen eröffnen sich neue Wege in die Freiheit. Die Technik hinter den Anwendungen gilt als neutral oder Zauberei und bleibt meistens intransparent.

Die Meinungsfreiheit hat sich durchs Digitale verändert, während sich die Zugänglichkeit verschiedener Medien erhöht hat. Dennoch sind es einzelne US-amerikanische Firmen, die bekannte Markt-Plattformen und Betriebssysteme bereitstellen. Dieselben diskutieren als Stakeholder auf europäischer Ebene und haben sich sogar im weltweiten Bildungssystem etabliert. Auch das US-amerikanische Reglement ist Veränderungen unterworfen. Keiner kann bestimmt sagen, ob und wie lange es freies Internet *by default* am US-amerikanischen Markt geben wird, während Friedensprojekte zunehmend unter Druck geraten.

Der Mangel an Kontrolle über den digitalen Raum zieht die Europäische Kommission immer mehr in die Pflicht für digitale Infrastruktur und Freiheitsrechte im Internet zu garantieren. Es wird eine Cloud für offene Wissenschaft geben. Die anderen Bereiche sind marktzentriert und verschlossen. Ob die EU ihre Chance zur Verteidigung europäischer Werte wahrnimmt und weltweit eine diverse und freie Internet-Gesellschaft, anstelle der USA, stärkt, kann zurzeit bezweifelt werden. Juristische Lösungen für technische Probleme, wie Uploadfilter, zeigen Fallen in der Internetregulation. Wirtschaftliche und kriminalistische Interessen dominieren die Regulierungsbedürfnisse der europäischen Politik. Das, obwohl es auch zivilgesellschaftliche Organisationen, wie European Digital Rights (EDRI) oder Wikimedia, mit an den Verhandlungstisch geschafft haben. Die Umsetzung der europäischen Digitalstrategie wird von Gesetzen und Abmachungen begleitet. Die Differenzierung verschiedener Daten nach Industriezweigen und Schwerpunkten kann zu mehr Datenreichtum, Kapital und Intelligenz führen. Datensicherheit wird damit nicht gewährleistet. Algorithmische Entscheidungen beeinflussen die Gesellschaft schon jetzt. Informierte Zustimmung wird erteilt, ohne die entsprechenden Datenschutzerklärungen und Nutzungsbedingungen gelesen zu haben. Internationale Kriminalität ist ein

ernstzunehmendes Feld. Bei mangelnder technischer Transparenz oder Kompetenz bleibt Kontrolle Illusion. (Pommerening 2013:26)

1.2.3.

Die Prüfung

Die institutionalisierte Prüfung ist schon länger bekannt als das Internet. Wir stellen den Alltag auf die Prüfung, um Täuschungen zu entgehen, um Fakten zu recherchieren, um die Welt zu begreifen. Das allein ist keine Wissenschaft, die als institutionalisierte Form der Prüfung gesehen werden kann. Foucault nähert sich der Prüfung aus einem sozialwissenschaftlichen Kontext, wenn er 1975 über *"Die Mittel der guten Abrichtung"* schreibt. (Foucault, 1977:220ff) Das Werk *"Überwachen und Strafen"* erschien erst letztes Jahr in seiner 17. Auflage. Ich habe es im Zuge der COVID-19 - Krise gelesen. Der theoretische Ansatz ist gut auf soziale Daten anwendbar: *"Die Prüfung macht auch die Individualität dokumentierbar."* (Foucault, 1977:243) *"So formieren sich eine Reihe von Codes der Disziplinarindividualität, mit denen sich die durch die Prüfung ermittelten individuellen Züge vereinheitlichen und verschlüsseln lassen..."* (Foucault, 1977:244)

Um *"Disziplinarindividualität"* zu verstehen, sei angefügt, was Foucault als Disziplinarmacht versteht: *"...die Disziplarmacht: sie setzt sich durch, indem sie sich unsichtbar macht, während sie den von ihr Unterworfenen die Sichtbarkeit aufzwingt. In der Disziplin sind es die Untertanen, die gesehen werden müssen, die im Scheinwerferlicht stehen, damit der Zugriff der Macht gesichert bleibt. Es ist gerade das ununterbrochene Gesehenwerden, das ständige Gesehenwerdenkönnen, ... was das Disziplinarindividuum in seiner Unterwerfung festhält. Und das Examen ist die Technik, durch welche die Macht (...) diese in einem Objektivierungsmechanismus einfängt."* (Foucault, 1977:241)

Als Beispiel nennt er Krankenhäuser: *"Die anderen Neuerungen der Disziplinarschrift betreffen die Korrelierung dieser Elemente, die Speicherung und Ordnung der Unterlagen, die Organisation von Vergleichsfeldern zum Zwecke der Klassifizierung, Kategorienbildung, Durchschnittsermittlung und Normenfixierung."* (Foucault, 1977:244f) ... *"Als rituelle und zugleich >wissenschaftliche< Fixierung der individuellen Unterschiede, als Festnagelung eines jeden auf seine eigene Einzelheit (...) zeigt die Prüfung das jeder seine eigene Individualität als Stand zugewiesen erhält, in der er auf die ihn charakterisierenden*

Eigenschaften, Maße, Abstände und >Noten< festgelegt wird, die aus ihm einen >Fall< machen.” (Foucault, 1977:247)

“In dieser winzigen Technik steckt nämlich ein ganzer Wissensraum und ebenso ein ganzer Machttyp. (...) Die Überlagerung der Machtverhältnisse und der Wissensbeziehungen erreicht in der Prüfung ihren sichtbarsten Ausdruck. (...) Sie errichtet über den Individuen eine Sichtbarkeit, in der man sie differenzierend behandelt.” (Foucault, 1977:238) Die Digitalisierung verschärft Foucaults Erkenntnisse spürbar.

Soziale Daten können Prüfungsergebnisse sein. Doch ich glaube nicht, dass sich Menschen der Prüfungssituation bewusst sind, besonders wenn sie Daten beiläufig produzieren und versenden. Niemand kann wirklich *das ganze Internet* überwachen. Technische Gegebenheiten spielen wenig Rolle in politischen Debatten. Man könnte womöglich eine Diskussion anschließen, die hinterfragt, was *das ganze Internet* bedeutet. Dafür ist hier kein Platz. Ich möchte aufzeigen, dass die Digitalisierung und die Prüfung ein Wissens- und Machtraum sind, in dem soziale Daten auftreten. *“... je anonym und funktioneller die Macht wird, um so mehr werden die dieser Macht Unterworfenen individualisiert: und zwar weniger durch Zeremonien als durch Überwachung...”* (Foucault, 1977:248) Ich meine, die Überwachung wird langsam durch Kontrolle ersetzt, die Daten retrospektiv verwendet. Mit den Ergebnissen wird die Zukunft bewertet, die man wiederum evaluiert. Die Prüfung behandelt alle gleich, Rassismen werden bei der standardisierten Erfassung der Realität ausgeblendet. Die Prüfungskriterien bestimmen die Prüfer_innen oder andere Zuständige.

1.3.

Definition

Soziale Daten sind Daten die sozial, erfassbar, wiederverwendbar, historisch bedingt und eine Summe aus personenbezogenen Daten darstellen, in dem Sinn, dass soziale Daten in ihrer Entstehung und durch ihren Inhalt auf einzelne Personen zurückführbar sind.

1.3.1.

Charakteristiken

Die folgenden Teilkapitel gehen näher auf die Definition ein.

1.3.1.1.

sozial

Die Natur des Sozialen ist, sofern es eine solche gibt, nicht eindeutig bestimmbar. Sozialwissenschaftler_innen kommen dennoch auf methodisch geleitete Erkenntnisse, die Auswirkungen auf die Gesellschaft haben können. Der quantitative Methodenkanon setzt sich verstärkt mit mathematischen Methoden auseinander, um soziale Daten vergleichbar und bewertbar zu machen. Qualitative Methoden interpretieren und erheben Daten, die sich nicht auf Mathematik verkürzen lassen. Sozialwissenschaftler_innen unterteilen das Soziale in verschiedene Bereiche, wie Wirtschaft, Kultur, Politik, Familie oder Religion. Durch die Digitalisierung sind soziale Daten aus all diesen Bereichen vorhanden, da sie von Menschen hergestellt werden. Soziale Daten sind ein sozialwissenschaftliches Phänomen. Sie sind nur existent, wenn sie Menschen erfassen.

1.3.1.2.

erfassbar

Soziale Daten können nur existieren, wenn sie auch erfassbar sind. Theoretisch ist ein Lächeln, ein Gedanke, eine Praxis erfassbar. Zeiterfassungssysteme machen Arbeitszeit messbar. Click-Verhalten auf Webseiten kann analysiert werden. Haushaltsgröße, Alter, Geschlecht, Herkunft, Haustiere, Religion, die sexuelle Orientierung, der Wohnort, Handynummern oder E-Mail-Adressen sind erfassbar. Auch der Stromverbrauch von Wohnungen, Autos und anderen "smart devices" sind auslesbar. Haarfarbe, Gesichtsgeometrie, Augenfarbe, Tattoos, Stimme, DNA-Spuren, Fingerabdrücke, Handschrift oder Gehtempo sind erfassbar. Kaum könnte man tatsächlich alle sozialen Daten erfassen. Ob

aus einem Ereignis ein *Fall* gemacht wird, wie Foucault ihn definiert, hängt von Prüfer_innen ab.

1.3.1.3.

wiederverwendbar

Soziale Daten können die ganze Lebensperiode genutzt werden, wenn kein technischer Prozess Wartezeit beansprucht. Ein Datensatz kann digital beliebig oft abgerufen und kopiert werden. Während Forscher_innen und Sicherheitspersonal früher persönlich über wertvolle Datensätze gewacht haben, halten nun technische Systeme vom Diebstahl ab. Während man schwere Akten nur umständlich entwenden kann, bleiben Datenlecks im Internet oft unentdeckt. Theoretisch kann ein Datensatz von Dieben und Forschenden gleichzeitig verwendet werden, ohne dass diese voneinander wissen. Während das Vernichten einzelner Akten einfach ist, wird das Löschen von Kopien im Digitalen quasi unmöglich, da man nicht wissen kann, ob und wo Kopien existieren. Soziale Daten können, sobald sie erfasst wurden, immer wieder verwendet und kopiert werden. Ob sie aufgrund des Alters, der Herkunft oder ihrer Qualität unverwertbar sind, hängt von den Menschen und von der Aufgabenstellung ab.

1.3.1.4.

historisch

Wie alle Phänomene im Sozialen sind auch soziale Daten historisch bedingt. Das Internet ist nicht aus dem Nichts entstanden und die Probleme, die sich darin spiegeln, zeigen eher soziale Wesen als technische Programmfehler. Internetregulation braucht es, seitdem man sieht, dass auch Kriminelle die freie Technik nutzen. Ob und welche sozialen Daten zu welchem Zweck erfasst werden, hängt auch von der Zeit ab. Die Geschichte ändert sich, genauso, wie sich Menschen ändern können. Soziale Daten sind nicht immer aktuell und damit nicht unbedingt richtig. Genau genommen sind sie nie aktuell, weil die Analyse retrospektiv, auf Grundlage bestehender Daten erfolgt. Vorhandene

soziale Daten wurden zu einem bestimmten Zeitpunkt, mit einer gewissen Technik, zu einem definierten Zweck, erfasst. Menschen, die bestimmte soziale Phänomene über einen längeren Zeitraum erforschen, beziehen sich auf soziale Daten, die zu historisch bestimmbar Zeitpunkten gemessen wurden. Ein Zweck könnte Forschung sein, die zu einem gewissen Zeitpunkt, unter bestimmten, zeitlich fluktuierenden Bedingungen, veröffentlicht wird. Naturwissenschaftliche Fakten, wie Bedingungen für Schwerelosigkeit, können ahistorisch verstanden werden. Soziale Daten verändern sich mit der Geschichte.

1.3.1.5.

Summe

Ich verwende soziale Daten im Plural. Singular wäre es *soziales Datum*, also ein Eintrag über eine Person oder Personengruppe. Soziale Daten sind eine Summe, also zumindest zwei soziale Daten. Die Summe sagt nichts darüber aus, welche sozialen Details erfasst, oder ob diese sinnvoll analysiert werden. Zudem wird nichts über die Methode, Herkunft oder die Zuverlässigkeit der Daten ausgesagt. Man kann Äpfel, Birnen und Bananen gemeinsam addieren, subtrahieren, oder andere statistische Verfahren anwenden. Ob das schlau ist, kommt auf die Aufgabenstellung an. Ich meine, dass soziale Daten bereits in Mengen vorhanden sind. Die Summe der sozialwissenschaftlichen Essays umfasst nicht die Zahl der darin verarbeiteten Daten. Auftragsforschung ist selten offen für die Wissenschaft. Die Kapazitäten zur Datenverarbeitung konnten technisch massiv gesteigert werden. Es gibt verschiedene Summen aus verschiedenen sozialen Daten. Die meisten Daten sind verschlossen und nicht nachvollziehbar.

Alles, was sich im Internet findet ist "user-generated-content", den man verarbeiten könnte. Prinzipien der Repräsentativität oder Zufallsstichproben rücken in den Hintergrund. Gesehen wird, was sichtbar ist. Sieht man nichts, kann auch das erfasst werden. Ob eine Summe sozialer Daten zur Prüfsumme

wird, um Menschen in Gruppen einzuteilen und welche Pläne für diese Gruppen gefasst werden, kann nicht allein durch die Summe bestimmt sein.

1.3.2.

Anwendungsfall "Big Social Data"

Olshannikova, E., Olsson, T., Huhtamäki, J. et al. stellen im Journal of Big Data 2017 ein Konzept zu "*Big Social Data*" vor. Der Essay ist online, open access und ohne Seitenangaben abrufbar. Mit dem Begriff "*Big Social Data*" beziehen sich die Autor_innen auf Daten, die in großen Mengen vorkommen: "*Big Social Data is any high-volume, high-velocity, high-variety and/or highly semantic data that is generated from technology-mediated social interactions and actions in digital realm, and which can be collected and analyzed to model social interactions and behavior.*" (Olshannikova, Olsson, Huhtamäki & Kärkkäinen, 2017)

Sie klassifizieren drei "*Big Social Data types*":

"1. Profile data It includes login data (usually a name/nickname/e-mail address with which other people identify the user); identity data (depends on the digital environment, i.e. for some services one should provide real first name and last name, mobile phone number, country, education, birthday); and personality data (e.g., profile pictures, tags of interest, slogan, personal signature in discussion forums) In many social media services, it is the personality data that the other users particularly focus on and analyze to assess the interestingness of the user.

2. Self-published content It incorporates publicly disclosed or socially restricted data (to trusted users or specific communities), such as most status updates in social media, pictures, videos, and other content that people add to services to represent themselves.

3. Data published by the community Self-representation could be complemented through person-related content shared by other users. This refers to collaboratively created pictures, narrations, videos, etc." (ebd.)

Die Definition von "*Big Social Data*", inklusive Klassifizierungen, ist nur auf digitale Daten anwendbar und richtet sich an die Disziplinen Social Computing, Big Data Science und Data Analytics, die gemeinsam "*Computational Social Science*" betreiben könnten. (Olshannikova et al., 2017 - Figur 2) Diese Perspektive scheint für ihr Feld anwendbar, doch

beschränkt es sich rein auf digitale und semantische Daten und vernachlässigt numerische Werte, die auch Soziales bedeuten können. Der Inhalt von Big Social Data, soziale Daten, wird im Essay bewusst vernachlässigt. Soziale Daten über soziale Daten, wie Forschungsergebnisse, also Aufzeichnungen von anderen, die Menschen persönlich betreffen, kommen im Konzept kaum vor. Big Social Data sind Daten die sozial, erfassbar, historisch, wiederverwendbar und die Summe aus personenbezogenen Daten darstellen - soziale Daten.

1.4.

Beispiele sozialer Daten

Das Motto "Private Daten schützen, öffentliche Daten nützen!", klingt einfacher, als es ist. "Soziale Daten nützen und schützen" liegt im Bestreben der Europäischen Union. Um das zu zeigen, werde ich soziale Daten, die ich in den analysierten europäischen Konzepten gefunden habe, *Schutz* und *Nutzen* zuordnen. Die Begriffe, die mit einem * gekennzeichnet sind, kommen nicht explizit in den europäischen Beispieltextrn des zweiten Teils vor und sind trotzdem erwähnenswert. Die Zuordnung verläuft fließend, doppelte Nennung ist möglich.

Schutz: sensible Daten, private Daten, *Dateneigentum, pseudonymisierte Daten, Gesundheitsdaten, personenbezogene Daten, Dateninhaber, citation data, Referenzdatensätze, Datenschutzvorschriften, Datenschutzvorkehrungen, Datenschutzmaßnahmen, Datenschutzgrundsätze, Datenschutzregeln, Europäischer Datenschutzausschuss, Datenschutzpflichten, Datenschutz-Aufsichtsbehörden, Datenschutzsiegel, Datenschutzzertifizierung, Datenschutzprüfung, Standarddatenschutzklausel, Datenschutzbeauftragte, Datenerzeuger, Datenzugriff, Datenverwendung, Datenraum, Datenzugang, Geschäftsdaten, Datennutzung, Daten-Governance, internationaler Datenverkehr, Datenkompetenz, Massendatenanalyse, datengetriebene Produkte, gemischte Datensätze, Datengenehmigungsbehörde, Dateninfrastruktur, Datenverarbeitungsinfrastruktur, Datenvorteil, Datenverarbeitungsdienste, Datennutzer, Datenverarbeitungsvorgänge, Datenethikkommission, Verbraucherdatenrecht, Datenübertragbarkeit, Datenexperten,

Datengesellschaft, Datenwertschöpfungskette, Datenökosysteme, Datenakkumulation, Datenrevolution, Datenprodukte, Transaktionsdaten, Datenqualität, Datenquellen, Datenübermittlung, Datenschutzniveau, Kontaktdaten, genetische Daten, RNS-Analyse-Daten, DNS-Analyse-Daten, biometrische Daten, Datenmissbrauch, Datenschutz-Folgeabschätzung, Profiling, Beschäftigtendaten, Datenminimierung, Standortdaten, daktyloskopische Daten, Daten zu sexuellen/politischen/ethischen/religiösen/gewerkschaftlichen Gesinnungen, Datenschutzpraxis, Speicherfristen (für soziale Daten), Datensicherheit, Datenschutzs Schulungen, *Polizeidatenbanken ...

Nutzen: Verhaltensdaten*, (soziale) Mikrodaten, Stromnetzdaten, Bankkontodaten, Finanzdaten, KFZ-Daten, Gesundheitsdaten, Berichtsdaten, Sozialdaten, pseudonymisierte Daten, (soziale) Metadaten, öffentliche (soziale) Daten, (soziale) Forschungsdaten, (soziale) agile Datenwirtschaft, Datenmengen, offene (soziale) Daten, nicht-personenbezogene (soziale) Daten nach Verordnung (EU) 2018/1807., Datenkompetenz, (soziale) Datenpools, Datenvorteil, IoT-Daten, Daten-Governance, Datenverarbeitungsmarkt, Datenströme, Datenübertragbarkeit, Datenwertschöpfungskette, Datenökosystem, Datenaltruismus, Datentechnologie, (soziale) Echtzeit-Daten, *Near-time-Daten, Datenverwalter, Gesundheitsdatenraum, Mobilitätsdatenraum, Finanzdatenraum, Energiedatenraum, Energieverbrauchsdaten, Kompetenzdatenraum, Datenverkehrsmessungen, (soziale) Fahrzeugdaten, Patientendaten, Bilddaten, Datenübermittlung, datenschutzfreundlich, Datensicherheitsrisiken, Profiling, Datenverarbeitungssituationen, aggregierte (soziale) Daten, besondere (soziale) Datenkategorien, *soziographische Daten, *demographische Daten, *user-generierte Daten, *Online-Daten, *Bevölkerungsdaten, *(sozialer) Datensatz, *Organisationsdaten, *Projektdate n, *Verwaltungsdaten ...

1.5.

Grenzfälle

In Datenkorrelationen, gerade bei unterschiedlicher Datenherkunft, können verschiedene Situationen auftreten, die soziale Daten für die Forschung non-existent machen.

1.5.1.

Datenverarbeitung

Technische Systeme können Daten speichern, müssen sie aber nicht. Soziale Daten werden in vielen Schritten verarbeitet, doch die wenigsten protokollieren sie. Ab und zu müssen Daten zwischenzeitlich gespeichert werden, um technische Aktionen ausführen zu können, beispielsweise im Rahmen einer Dienstleistung. Das bedeutet aber nicht, dass Datenverarbeiter_innen die Daten längerfristig speichern, auslesen oder prüfen. Werden soziale Daten ausschließlich maschinell verarbeitet, ohne den Zweck Daten auszulesen, sind faktisch keine sozialen Daten vorhanden. Sobald soziale Daten aus Datenverarbeitungsprozessen gespeichert und ausgelesen werden, sind sie als solche zu erkennen.

Beispiel Bewegungsstromanalyse: Über Netzwerk-Anbieter kann die Verfolgung einzelner Rufnummern erfolgen. Aggregierte Bewegungsdaten, die Gruppenbewegungen von mehreren Personen auf einer Karte anzeigen, sind ein Ergebnis dieser sozialen Daten, die aus Datenschutzgründen verschlossen bleiben. Die Bewegungskarte zeigt andere, aggregierte soziale Daten, die für Forschung verwendbar sind.

1.5.2.

Fakten

Anders als bei naturwissenschaftlichen Fakten, ändern sich Menschen und damit auch die sozialen Daten über sie. Personenbezogene Daten können fehlerhaft erhoben oder schlichtweg falsch sein. Werden Daten von Testpersonen selbst generiert, kann man nicht nachvollziehen, ob wahre oder falsche Aussagen hinter dem Antwortverhalten stecken. Durch unterschiedliche Perspektiven werden Ereignisse verschieden bewertet, oft ohne die Intentionen der Datenspender_innen zu bedenken. Werden Daten von anderen erfasst, kann kaum festgestellt werden, ob die Fremdwahrnehmung mit der Eigenwahrnehmung der Beforschten übereinstimmt. Soziale Daten sind keine neutralen oder ahistorischen Fakten.

Beispiel Social-Media: Forschung im digitalen Bereich ist spannend und maschinenlesbar. Kommentare und Likes entstehen aus den unterschiedlichsten Gründen und Debatten sind

online abgebildet. Tatsache ist, soziale Daten existieren auch online. Daraus kann man kaum schließen, dass Online-Daten die faktischen Realitäten einer Gesellschaft abbilden. Menschliches Verhalten ändert sich über die Zeit, auch online.

1.5.3.

Private Daten

Private Daten sind per Definition sozial, dennoch per se non-existent für die Forschung. Im kommenden Abschnitt wird die Datenschutzgrundverordnung näher analysiert, die eine rechtliche Grundlage für Privatsphäre im Digitalen schafft. Das Aufkommen von Big Data wirft Fragen auf, ob bisherige Pseudonymisierungsverfahren ausreichen, wenn sich diverse Daten verschiedener Herkunft zu einer Person zusammenfügen lassen. (Rocher, Hendrickx & de Montjoye, 2019) Jeder Mensch darf Daten über sich preisgeben, auch öffentlich und online. Die Privatsphäre natürlicher Personen bleibt trotzdem ein besonders geschütztes Menschenrecht.

Beispiel Gesundheitsdaten: Zur Pandemie-Bekämpfung ist beinahe alles erlaubt, solange die Maßnahmen verhältnismäßig sind. Darunter auch die Aufzeichnung von Gesundheitsdaten, wenn das erforderlich und zweckmäßig scheint. Diese Daten gehören zu den sensibelsten Daten, die ein Mensch besitzt. Trotzdem braucht es keine informierte Zustimmung natürlicher Personen um Gesundheitsdaten online verarbeiten zu dürfen, solange es der öffentlichen Gesundheit zuträglich scheint. Weniger noch, die elektronische Gesundheitsakte, Daten der Krankenkassen und Anwendungen der individuellen Medizin werden von Institutionen über Privatpersonen erstellt. Menschen werden nicht gefragt, ob sie ihre Daten digitalisieren wollen...

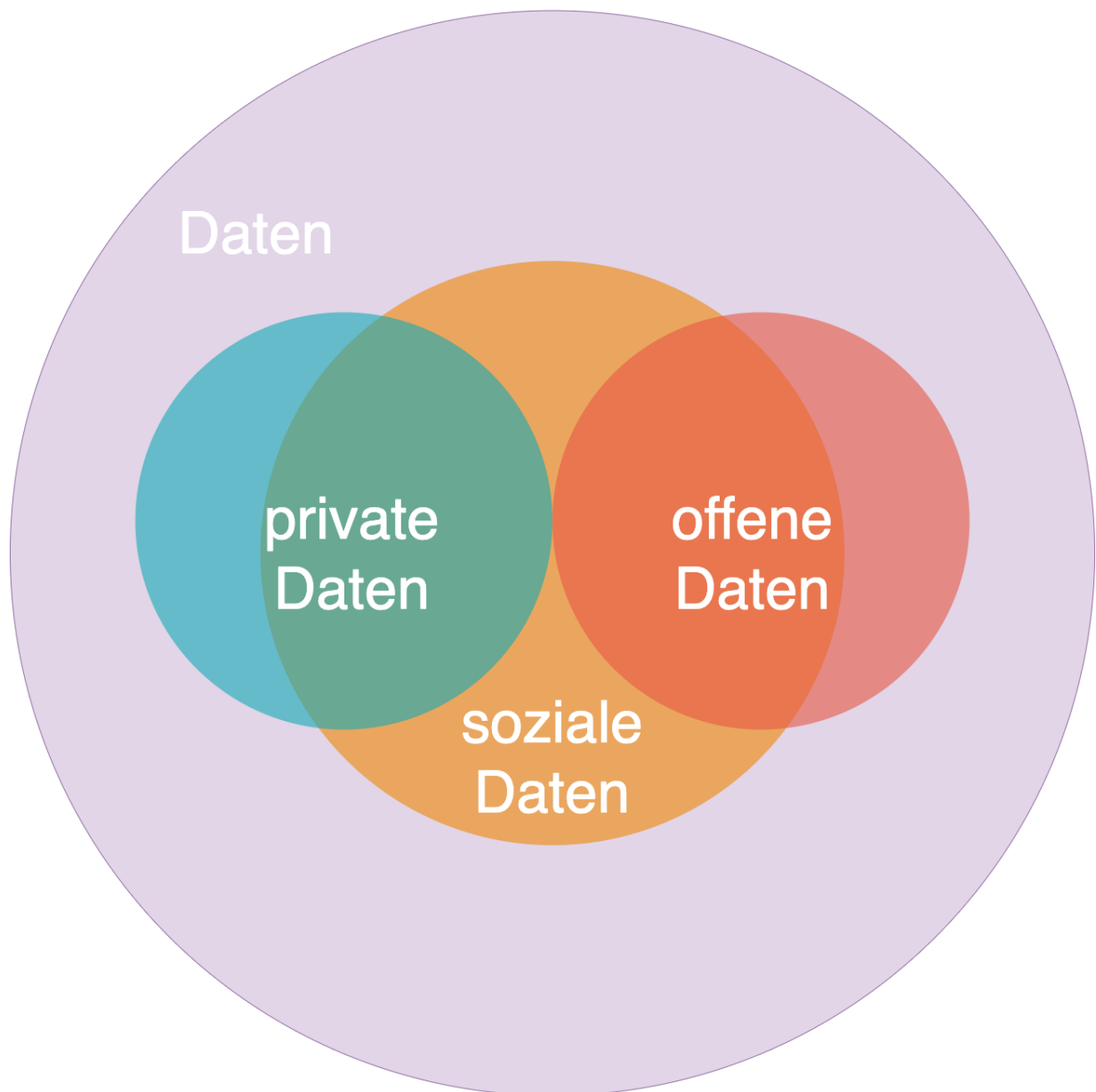


Abbildung 1: Soziale Daten

2.

Soziale Daten in der EU

Soziale Daten können als epistemologische Grundlage jeder Metatheorie verstanden werden die das Soziale berührt. Sie lässt sich auf europäische Policies anwenden. Es gibt zahlreiche europäische Institutionen, die sich mit sozialen Daten befassen, wie Regelungen im E-Commerce, die Strategie zur Bekämpfung von Desinformation oder Open Science Projekte. Dazu kommen nationale Gesetze und Normen. Drei europäische Projekte, die DSGVO, Zenodo und die Europäische Datenstrategie, befassen sich auf verschiedenen Ebenen mit sozialen Daten. Die Beispiele zeigen ein diverses Bild im Umgang mit sozialen Daten, während sie diese gleichsam konstituieren.

2.1.

Die Datenschutz-Grundverordnung

Das 88 Seiten lange Dokument wurde 2016 verabschiedet und ist gültiges Recht in Europa. Die DSGVO ("Datenschutz-Grundverordnung") ist für alle frei zugänglich und einsehbar. (Amtsblatt der Europäischen Union, 2016) Ungefähr die Hälfte des vorliegenden Amtsblatts erklärt Erwägungen, die den Rechtstext begründen und näher ausführen. Der größere Teil des Dokuments besteht aus dem Gesetzestext der *Verordnung*. Ich habe das deutsche Dokument analysiert. Andere gültige Normen, wie Cookie-Banner oder Gesetze, berufen sich auf die Datenschutz-Grundverordnung. Die meisten Beispiele sozialer Daten aus dem ersten Teil, die sich auf Datenschutz beziehen, kommen in diesem Amtsblatt vor. Diese Verordnung konstituiert Privatsphäre im digitalen Raum. Pseudonymisierte Daten bleiben personenbezogene Daten, auch wenn sie für die Wissenschaft genutzt werden. Soziale Daten sind Mittelpunkt der Verordnung.

In elf Kapiteln und 99 Artikeln schafft man Institutionen, Zuständigkeiten und Regeln für personenbezogene Daten, wie Standarddatenschutzklauseln (Art. 46, 58), Datenschutzsiegel

(Art. 42, 43, 57, 70) oder den Grundsatz der Datenminimierung (Art. 5, 25, 47, 89). Ich analysiere den Text semantisch, nicht juristisch und zeige damit soziale Daten in Europa.

Alle Lesenden haben eine persönliche Vorstellung was Standarddatenschutzklauseln sein könnten. Webanbieter, die soziale Daten Dritter nutzen, agieren durch die Verordnung verantwortlicher und transparenter. Man sieht viele Standarddatenschutzklauseln, die einem den Schutz der persönlichen Daten zusichern. Nicht alle Standarddatenschutzklauseln müssen im Zusammenhang mit der DSGVO stehen. Im Gesetzestext bezieht man sich konkreter auf Standarddatenschutzklauseln, die von Aufsichtsbehörden festgelegt oder genehmigt werden können. Diese werden besonders beim Datenverkehr mit Drittländern relevant, wenn es darum geht, europäische Standards zu halten. Das Datenschutzsiegel wird vom Europäischen Datenschutzausschuss ausgearbeitet, damit genau bestimmte Zertifizierungsstellen bestehenden Datenschutz bei Produkten bescheinigen können. Das Prüfzeichen garantiert, dass bestimmte DSGVO-relevante Kriterien eingehalten sind. Derzeit liegen keine empirischen Befunde vor, die darauf hinweisen, dass sich das freiwillige Siegel durchsetzt.

Ähnlich steht es um den Grundsatz der *Datenminimierung*, der sich u.a. in Artikel 5 (1/c) findet: Personenbezogene Daten müssen *“dem Zweck angemessen und erheblich sowie auf das für die Zwecke der Verarbeitung notwendige Maß beschränkt sein.”* Welche sozialen Daten für einen Zweck angemessen und erheblich sind, ist durch die DSGVO höchstens vage zu vermuten und es gibt viel Interpretationsspielraum, solange Verantwortliche ihr Verhalten offenlegen. Mobilitätsdaten, die einen ganzen Datenraum in der europäischen Datenstrategie darstellen, können als Beispiel dienen: Welche sozialen Daten benötigen Autohersteller über ihre Kund_innen? Welche sozialen Daten braucht der Bahn-, Schiff- und Flugverkehr über Fahrgäste? Datenerhebungen können gesetzlich vorgeschrieben werden, um Personen besser verfolgbar zu machen. Bis zum Jahr 2025 werden 80% der Daten im Bereich *Internet of Things* anfallen. (Europäische Kommission, 2020:2) Darunter versteht man vernetzte Objekte wie Autos oder Haushaltsgeräte. Ob all diese Daten notwendig sind, kann ich kaum beurteilen. Da zumindest 2% der aktuellen CO2 Emissionen schon jetzt aufs Internet zurückzuführen sind, könnte es aus meiner Sicht nicht schaden, den Grundsatz der Datenminimierung vehementer zu vertreten.

In den Artikeln 49 und 89 wird festgelegt, dass es Ausnahmen für soziale Daten gibt. Mitgliedstaaten dürfen, laut Art. 62, keine anderen Mitgliedstaaten wegen Datenschutzdelikten klagen. Kapitel IX behandelt besondere Verarbeitungssituationen. Artikel 85 (1): *“Die Mitgliedstaaten bringen durch Rechtsvorschriften das Recht auf den Schutz personenbezogener Daten gemäß dieser Verordnung mit dem Recht auf freie Meinungsäußerung und Informationsfreiheit, einschließlich der Verarbeitung zu journalistischen Zwecken und zu wissenschaftlichen, künstlerischen oder literarischen Zwecken, in Einklang.”* Europäische Behörden, Wissenschaft und Kunst können freier agieren, wobei man davon ausgehen muss, dass diese Bereiche Privatsphäre als Menschenrecht achten.

Der *Europäische Datenschutzausschuss* (Art. 65, 94 und Kapitel VII Abschnitt 3), *Datenschutzbeauftragte* (Art. 13, 14, 30, 33, 35, 36, 47, 57, 68, 75 und Kapitel IV Abschnitt 4) und die *Europäischen Aufsichtsbehörden* (Art. 4, 12 - 15, 27, 28, 30, 31, 33, 34, 35 - 37, 39, 40 - 43, 46, 47, 49, 50, 60, 61 - 68, 70, 74, 75, 77, 78 - 80, 83, 85, 90, 91, 97 und Kapitel VI) fungieren als Entscheidungsträger. Wobei die Hauptverantwortung beim *Verantwortlichen*, also bei den datenverarbeitenden Personen oder Gruppen, bleibt. Die Verantwortlichen müssen auch Privatpersonen offenlegen, welche sozialen Daten über sie verarbeitet werden (Art. 15).

2.2.

Europäische Datenstrategie

Die Europäische Datenstrategie fokussiert sich, im Vergleich zur DSGVO, nicht auf den Schutz sozialer Daten, sondern auf die Nutzbarmachung des europäischen Datenreichtums. (Europäische Kommission, 2020) Darunter fallen allerlei Daten aus verschiedenen wirtschaftlichen Sektoren. Diese müssen nicht unbedingt sozial sein. Maschinelle und naturwissenschaftliche Daten sind auch ohne Personenbezug nutzbar. Beispiele sozialer Daten aus der Europäischen Strategie finden sich im Abschnitt 1.4. . Der überwiegende Teil der nutzbringenden Daten aus Abschnitt 1 sind in diesem Dokument zu finden.

Die Strategie ist kein Rechtstext, sondern zeigt Pläne der Europäischen Kommission im Umgang mit europäischen Daten. Ein Rechtstext wird angestrebt, wenn Abmachungen wie

Verhaltenskodexe (*Codes of Conduct*) mit Stakeholdern unzureichend sind. Das Strategiepapier selbst ist nicht im luftleeren Raum entstanden, sondern bildet Wünsche und Interessen von Stakeholder_innen ab. Als frei zugängliche, öffentliche, europäische Kommunikation informiert die Datenstrategie interessierte Bürger_innen. Man könnte argumentieren, dass der politische Prozess über derartige Kommunikationspapiere transparent gemacht wird. Für mich bleibt die Datenstrategie und ihre Umsetzung eher undurchsichtig, da große Teile der entsprechenden Kommunikation unzugänglich sind. Ich würde eine Veröffentlichung aller vereinbarten Codes of Conduct (o.ä.) vorschlagen. So wären gültige Abmachungen transparent wie europäische Gesetze und alle Stakeholder, sowie europäische Bürger_innen könnten gleichermaßen profitieren.

Das untersuchte Strategiepapier fasst europäische Entwicklungen auf 40 Seiten zusammen. Pläne und Wünsche werden damit gestärkt. Im ersten Teil der Datenstrategie - Kapitel 1-4 - werden Intentionen, Probleme und Ziele definiert, um anschließend, ab Kapitel fünf, eine Strategie zu präsentieren, die diesen Entwicklungen begegnet. Nach einem kurzen Fazit folgen zehn Seiten Anhang, der *„zusätzliche Hintergrundinformationen über die sektorspezifischen Maßnahmen und Rechtsvorschriften...“* darstellt. (Europäische Kommission, 2020:30). Darin werden die Datenräume aus der Strategie näher ausformuliert. Durch die Datenstrategie soll die EU in die Lage versetzt werden, *„zur attraktivsten, sichersten und dynamischsten datenagilen Wirtschaft der Welt zu werden, damit Europa mithilfe von Daten bessere Entscheidungen treffen und das Leben aller Bürgerinnen und Bürger verbessern kann.“* (Europäische Kommission, 2020:29)

Die Datenstrategie beruht auf vier Säulen: Erstens, einem sektorenübergreifenden Governance-Rahmen, zweitens, Investitionen in Daten, Interoperabilität und Infrastruktur, drittens, Handlungskompetenzen des Einzelnen und Investitionen in KMUs und viertens strategische Sektoren und Bereiche, die von öffentlichem Interesse sind. (Europäische Kommission, 2020:14ff) Zu jeder Säule werden konkrete Maßnahmen und entsprechende Zeitpläne angeführt. (ebd.) Die EU verfolgt einen offenen, aber proaktiven internationalen Ansatz, *„in dem die wirtschaftliche und technologische Entwicklung unter uneingeschränkter Einhaltung des EU-Rechts voranschreiten kann.“* (Europäische Kommission, 2020:27)

Aus meiner Sicht gibt es bei der Durchsetzung und Gewährleistung von Privatsphäre-Rechten Handlungsbedarf. Vielleicht kann die Datenstrategie auch als die geplante Evaluation der DSGVO gesehen werden, die bei der Erstellung dieser Arbeit noch nicht vorliegt. Folgendes Zitat zeigt, dass unsere Grundwerte international ungeschützt sein könnten: *“Die Kommission ist davon überzeugt, dass die internationale Zusammenarbeit auf einem Ansatz beruhen muss, der die Grundwerte der EU, einschließlich des Schutzes der Privatsphäre, fördert.”* (Europäische Kommission, 2020:28) Soziale Daten sollen geschützt bleiben, selbst wenn sie sich außerhalb der EU befinden. (ebd.) Damit reagiert man auch auf zahlreiche internationale Dienstleister. Potenziell fallen derart viele Daten zu jeder Person an, dass Pseudonymisierungsverfahren durch erhältliche Daten neutralisiert werden könnten. Jeder Einzelne kann durch soziale Daten sichtbar werden. Durch freies Internet und technologischen Fortschritt hat niemand Überblick, welche sozialen Daten im Umlauf sind. Europäischen Bürger_innen soll *dezentrale Technologie* an die Hand gegeben werden, um Kontrolle über die eigenen sozialen Daten zu behalten. (Europäische Kommission, 2020:12) Das könnte zumindest die Datenqualität und Richtigkeit einzelner Datensätze verbessern.

Die vierte Säule der europäischen Strategie zeigt zehn europäische Sektoren, die gemeinsame Datenräume bilden, um Daten effizient zu nutzen. (Europäische Kommission, 2020:26f) In allen zehn Datenräumen fallen soziale Daten an. Im *Mobilitätsdatenraum*, dem *Gesundheitsdatenraum*, dem *Finanzdatenraum*, dem *Energiedatenraum*, dem *Kompetenzdatenraum*, im *Datenraum für die öffentliche Verwaltung* und in der *Cloud für offene Wissenschaft* stehen soziale Daten im Mittelpunkt. Die genannten Datenräume unterscheiden sich zudem vom *Industriedatenraum*, dem *Agrardatenraum* und dem *Datenraum für den europäischen Grünen Deal*, weil dort andere Daten im Mittelpunkt stehen. Auch wenn maschinelle Berechnungen durch Big Data nützlich sein können, bedeutet das nicht unbedingt, dass die europäische Bevölkerung Datenexperimente mit ihren Gesundheitsdaten gutheißt. Die höchsten Sicherheitsstandards werden bei allen europäischen Online-Clouds geboten. Absolute Geheimhaltung kann es im Netz nicht geben, nur gute Sicherheitsvorkehrungen. Informierte Zustimmung der Testpersonen ist Qualitätsanforderung jeder guten Wissenschaft, die hoffentlich für alle - europäischen - Clouds gilt.

2.3.

Zenodo

Die digitale Infrastruktur Zenodo bietet eine andere Art Raum für soziale Daten. Es handelt sich um die größte Infrastruktur für offene Wissenschaft. Der Name ist bewusst an Zenodotus angelehnt. (Zenodo, <https://about.zenodo.org/>) Er war der erste Bibliothekar in der antiken Bibliothek von Alexandria. Zudem nutzte er nachweislich Metadaten (ebd.). Physisch findet man Zenodo in der Schweiz, bereitgestellt vom CERN. Das Projekt wird von der Europäischen Union mitfinanziert. (Zenodo, <https://about.zenodo.org/infrastructure/>) Es ist der wichtigste europäische Online-Speicher für Wissenschaftler_innen, die offen publizieren, da Wissen strukturiert und langfristig aufbewahrt werden kann.

Zenodo verfolgt einen anderen Ansatz als Verlagshäuser, die publizierende Wissenschaftler_innen für *Open Access* bezahlen lassen. Die Infrastruktur ist eher eine partizipative, kostenlose, digitale Bibliothek, die Wissenschaftler_innen dabei hilft, offener und nachhaltiger zu arbeiten. Wissen, Kunst und Metadaten können selbstbestimmt veröffentlicht, aktualisiert und gelöscht werden. Arbeiten werden strukturiert zugänglich gemacht. Zenodo bekennt sich zu den *FAIR-Prinzipien*, die auch in der Europäischen Datenstrategie erwähnt werden. (Zenodo, <https://about.zenodo.org/principles/>)

Zenodo fördert offene Wissenschaft ohne Grenzen, indem es weltweit für alle zugänglich ist. Nutzer_innen müssen keine europäischen Wissenschaftler_innen sein. Alle können die Plattform kostenlos nutzen. Zudem ist Zenodo *open source*. (CERN, <https://github.com/zenodo/zenodo>) Das bedeutet, es steht allen frei den Quellcode zu nutzen. Alle können Verbesserungsvorschläge einbringen oder Schwachstellen melden. Nutzer_innen können nachvollziehen wie Zenodo als Programm funktioniert. Das stärkt Offenheit und Transparenz. So gewinnt Zenodo das Vertrauen von Wissenschaftler_innen, deren professioneller Reichtum oftmals von sicher verwahrten Daten abhängt.

Nicht alle gespeicherten Informationen müssen öffentlich gemacht werden. Sensible Datensammlungen können beispielsweise als referenzierbare Primärquelle abgespeichert werden, ohne dass die Daten damit veröffentlicht sind. So können mehr detaillierte und strukturierte Forschungsdaten gesammelt werden, die in einem wissenschaftlichen Essay

kaum Platz finden. Dateninhaber_innen können spezielle Konditionen festlegen, unter denen sie Zugriff auf ihre Daten gewähren, solange sie Zenodo ohne unmittelbare Gewinnabsicht nutzen.

Alle Inhalte auf der Plattform sind adäquat lizenziert und verfügen über eine *eindeutige digitale Kennzeichnung (Digital Object Identifier - DOI)*. Damit sind die Daten sozial, auch wenn man aus einzelnen Arbeiten Fakten generieren kann. Die eindeutige Kennzeichnung digitaler Objekte erlaubt präzises Zitieren. Alle Inhalte auf Zenodo sind online, von allen Geräten aus, abrufbar. Die Lizenzierung beugt Rechtsunsicherheiten vor. Wissenschaftler_innen können Arbeiten besser nachvollziehen, testen und validieren, weil die Datengrundlage vorhanden ist.

Daten und Arbeiten können aktualisiert werden. Die Unmittelbarkeit des Digitalen gibt Wissenschaftler_innen mehr Freiheit, Arbeiten zu verbessern. Neuere Versionen eines Inhalts können problemlos hochgeladen werden, wobei jede Version eine *eindeutige digitale Kennzeichnung (DOI)* besitzt, sodass ein Bearbeitungsverlauf sichtbar wird. Nutzer_innen werden automatisch auf die aktuelle Version eines Inhalts verwiesen. Zenodo schafft damit eine neue Möglichkeit, Fehler und Unklarheiten zu korrigieren, indem es Nutzer_innen dazu anleitet, Inhalte auch nach der Veröffentlichung zu überarbeiten. Feedback und neue Erkenntnisse können in einer neuen Version miteinbezogen werden. So kann ein und dieselbe Arbeit zu einem Thema auch über mehrere Jahre aktuell gehalten werden. Das ist bei klassischen Publikationen nicht möglich. Die Verknüpfung von aktuellem Wissen erreicht damit ein neues Level, welches offline nicht denkbar wäre.

Anders als die DSGVO oder die europäische Datenstrategie hat Zenodo, als Plattform für soziale Daten, globale Relevanz. Die Nutzungsbedingungen von Zenodo sind keine rechtlichen Verordnungen oder Strategien, die sich auf ein geographisches Hoheitsgebiet beziehen. Die Plattform sammelt oder schützt keine Daten um strategische Vorteile zu ziehen. Alle Nutzer_innen profitieren gleich. Durch die Qualität der Plattform finden sich freiwillige Nutzer_innen, die sie gestalten. Plattformspezifische Regeln sind für alle gültig und einsehbar, egal von welchem Teil der Welt das europäische Angebot genutzt wird. Freie Wissenschaft als Zweck generiert soziale Daten durch Nutzer_innen, *user-generated-content*. Die Daten sind sozial, historisch, erfasst, wiederverwendbar und auf zumindest ein Personenprofil zurückverfolgbar. Dateninhaber_innen übernehmen

Rechte und Pflichten für ihre Publikationen, egal ob es sich dabei um besonders geschützte, soziale oder andere Analysedaten handelt.

2.4.

Zusammenfassung

Die Arbeit befasst sich mit sozialen Daten in der EU. Soziale Daten werden interdisziplinär zur Anwendung gebracht und kommen in wissenschaftlichen Texten vor. Im ersten Teil verorte ich soziale Daten als Grundlage der Sozialwissenschaften, der Digitalisierung und der Prüfung. Die Definition von *“Big social Data”* reicht nicht aus, um das Phänomen sozialer Daten zu verstehen. Um den Ursprung der Daten zu unterstreichen, zeige ich fünf Charakteristika sozialer Daten. Sie sind erfassbar, wiederverwendbar, historisch bedingt, auf Menschen bezogen und in ihrer Summe nutzbar.

Es gibt Grenzfälle, die das Bestehen sozialer Daten einschränken. In Europa sind soziale Daten durch diverse Normen omnipräsent. In der Arbeit greife ich drei wesentliche Beispiele heraus: Die Datenschutz-Grundverordnung, die Datenstrategie und die Infrastruktur Zenodo. Durch sie zeige ich, wie sich soziale Daten in der EU konstituieren.

Das Papier versteht sich als Theorie, die soziale Daten erforschbarer machen soll. Zahlreiche Fragen, die sich aus der Arbeit ergeben, bleiben offen: Inwiefern kann ein Individuum in einer Menge sozialer Daten geschützt bleiben? Was sind Grenzen der Pseudonymisierungsverfahren? Sind soziale Daten per se Gemeingut, weil sie die Summe von individuellen Besitztümern darstellen? Welche sozialen Daten müssen Bürger_innen bekanntgeben und in welchem Ausmaß behalten sie die Kontrolle darüber? Wie kann die neue Menge an sozialen Daten für die Sozialwissenschaften nutzbar werden? Sind Dateneigentümer_innen am Daten- und Wissensreichtum beteiligt, der durch sie generiert wird?

Diese und weitere Fragen könnten in zukünftiger Forschung geklärt werden.

Literaturverzeichnis

Amtsblatt der Europäischen Union (2016, 4. May): *VERORDNUNG (EU) 2016/679 DES*

EUROPÄISCHEN PARLAMENTS UND DES RATES vom 27. April 2016 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten, zum freien Datenverkehr und zur Aufhebung der Richtlinie 95/46/EG (Datenschutz-Grundverordnung). Online

verfügbar unter: URL:

<https://eur-lex.europa.eu/legal-content/DE/TXT/?uri=celex%3A32016R0679>

abgerufen am 25.7.2020

Anzinger, Heribert M, Hamacher, Kay, Katzenbeisser, Stefan (Eds.) (2013): *Schutz genetischer, medizinischer und sozialer Daten als multidisziplinäre Aufgabe.* Springer Berlin

Heidelberg Online verfügbar unter: URL:

<http://link.springer.com/10.1007/978-3-642-34741-2> [2013] abgerufen am 25.7.2020

Atzmueller, Martin (2012): Mining Social Media. *Informatik-Spektrum*, 35 (2) Online verfügbar

unter: URL: <http://link.springer.com/10.1007/s00287-012-0601-z> [2012-4-1] abgerufen

am 25.7.2020

CERN (n.d.): *Zenodo Github.* Online verfügbar unter: URL:

<https://github.com/zenodo/zenodo> abgerufen am 25.7.2020

DeMaris, Alfred (2004): *Regression with Social Data: Modeling Continuous and Limited Response Variables.* John Wiley & Sons, Inc. Online verfügbar unter: URL:

<http://doi.wiley.com/10.1002/0471677566> [2004-8-27] abgerufen am 25.7.2020

Europäische Kommission (2020, 19. February): *Eine europäische Datenstrategie.* Online

verfügbar unter: URL:

<https://ec.europa.eu/info/sites/info/files/communication-european-strategy-data-19>

feb2020_de.pdf abgerufen am 25.7.2020

Foucault, Michel (1977): *Überwachen und Strafen - Die Geburt des Gefängnisses*. (16.

Auflage.). Frankfurt am Main: Surkamp Verlag

Gross, Ana (2011): The Economy of Social Data: Exploring Research Ethics as Device. *The Sociological Review*, 59 (2_suppl) Online verfügbar unter: URL:

<http://journals.sagepub.com/doi/10.1111/j.1467-954X.2012.02055.x> [2011-12-1]

abgerufen am 25.7.2020

Guo, Lei, Vargo, Chris J, Pan, Zixuan, Ding, Weicon, Ishwar, Prakash (2016): Big Social Data

Analytics in Journalism and Mass Communication: Comparing Dictionary-Based Text Analysis and Unsupervised Topic Modeling. *Journalism & Mass Communication Quarterly*, 93 (2) Online verfügbar unter: URL:

<http://journals.sagepub.com/doi/10.1177/1077699016639231> [2016-6-1] abgerufen am

25.7.2020

Karimi, Azadeh, Tulloch, Ayesha I. T, Brown, Greg, Hockings, Marc (2017): Understanding the

effects of different social data on selecting priority conservation areas. *Conservation biology: the journal of the Society for Conservation Biology*, 31 (6), pp. 1439–1449 Online verfügbar unter: URL: <http://www.ncbi.nlm.nih.gov/pubmed/28425128> [2017-12-31]

abgerufen am 25.7.2020

Mitchell, Tom M (2009): Computer science. Mining our reality. *Science (New York, N.Y.)*, 326

(5960), pp. 1644–1645 Online verfügbar unter: URL:

<http://www.ncbi.nlm.nih.gov/pubmed/20019279> [2009-12-18] abgerufen am 25.7.2020

Olshannikova, Ekaterina, Olsson, Thomas, Huhtamäki, Jukka, Kärkkäinen, Hannu (2017):

Conceptualizing Big Social Data. *Journal of Big Data*, 4 (1) Online verfügbar unter: URL:

<http://journalofbigdata.springeropen.com/articles/10.1186/s40537-017-0063-x>

[2017-12-1] abgerufen am 25.7.2020

Rocher, Luc, Hendrickx, Julien M, de Montjoye, Yves-Alexandre (2019): Estimating the success of re-identifications in incomplete datasets using generative models. *Nature communications*, 10 (1), p. 3069 Online verfügbar unter: URL:

<http://www.ncbi.nlm.nih.gov/pubmed/31337762> [2019-7-23] abgerufen am 25.7.2020

Zenodo: *About Zenodo*. Online verfügbar unter: URL: <https://about.zenodo.org/> abgerufen am 25.7.2020

Zenodo: *Infrastructure*. Online verfügbar unter: URL:

<https://about.zenodo.org/infrastructure/> abgerufen am 25.7.2020

Zenodo: *Principles*. Online verfügbar unter: URL: <https://about.zenodo.org/principles/> abgerufen am 25.7.2020

Abbildung 1: Soziale Daten (Chlastak Maria, 2020 S.19)