UNIKASSEL
VERSITÄT

SoSe 2024

FB Wirtschaftswissenschaften

Empirische Forschungsmethoden

Hausarbeit zum Thema

Das Risiko der Re-Identifikation von anonymisierten Daten

Dozent: Prof. Dr. Markus Seyfried

Datum der Abgabe: 27.09.2024

Wörter: 3731

Verfasser: Benjamin Bleske

Studiengruppe: MPA 53B NRW

Matrikelnummer: 36104066

INHALTSVERZEICHNIS

Abbildungsverzeichnis

1 Einleitung

Spätestens seit 1890 wird der konzeptuelle Dreiklang aus Privatsphärenschutz des Individuums, gesellschaftlichem Wandel und der technischen Entwicklung miteinander kombiniert: Warren und Brandeis schufen in *The Right to Privacy* den Ausgangspunkt des heutigen individual-Datenschutzrechts (Warren/Brandeis 1890: 193,195). Mit dem Aufkommen von Sofortbildkameras und dem boomenden Zeitungsgeschäft begann eine Entwicklung (ebd.), die rund 100 Jahre später zum Inkrafttreten des ersten nationalen Bundesdatenschutzgesetzes (Bundesgesetzblatt 1977: 201ff) sowie zum ersten internationalen Vertragswerk führte, das auf den Schutz personenbezogener Daten ausgerichtet war: die *Konvention 108* des Europarats (Counsil of Europe 1981).

Zwei Jahr später konstatiert das Bundesverfassungsgericht im Volkszählungsurteil das Recht auf informationelle Selbstbestimmung (1983). Hierbei wurde dem Individuum die Befugnis erteilt, im Kontext der modernen Datenverarbeitung, sofern kein überwiegendes Allgemeininteresse vorliegt, höchstpersönlich über die Preisgabe und Verwendung seiner persönlichen Daten zu bestimmen (ebd.).

Auch heutzutage steht das Datenschutzrecht mit der Fortentwicklung von Technik, Gesellschaft, Wissenschaft und Wirtschaft in einem Spannungsverhältnis. Dazu wird den jeweils handelnden Akteuren, sobald sie personenbezogene Daten verarbeiten, das datenschützende Korsett der Datenschutz-Grundverordnung angelegt.

In der Vergangenheit hat hierbei die Anonymisierung von personenbezogenen Daten dabei geholfen, die enge Schnürung zu lockern. Damit konstatiert die Datenschutz-Grundverordnung im Erwägungsgrund 26 auch das, was bereits seit 2003 im § 3 Absatz 6 a.F. des Bundesdatenschutzgesetzes gegolten hat (Bundesgesetzblatt 2003: 66ff.): anonymisierte Daten sind keine personenbezogenen und daher keine schützenswerten Daten.

Diese „[...] legale Möglichkeit zur *Flucht aus dem Datenschutzrecht*." (Pohle/Hölzel 2020: 4) führt unweigerlich zur Ansammlung von großen Datenbeständen bei Organisationen, die je nach Kontext unterschiedlich miteinander kombiniert werden. Auch die Wissenschaft sammelt anonymisierte Daten, um diese in anderen Kontexten neu zu verknüpfen oder weiterzugeben, um die Nachnutzbarkeit von Primärdaten für Forschung und Lehre zu gewähr-

leisten (Huschka/Oellers 2013: 9f.). Das Problem ist jedoch, dass mit jedem weiteren verknüpften Datenbestand das Risiko der Re-Identifikation von Personen in Datenbeständen steigt (Deutscher Ethikrat 2017: 86ff.; Appenzeller/Orak 2024: 283).

Die Relevanz des Themas wird bei der Meldung über die Weitergabe von 73 Millionen pseudonymisierten Datensätzen zur Krankengeschichte der gesetzlich Versicherten in Deutschland an das Forschungsdatenzentrum des Bundes unterstrichen: im Kontext des im Jahr 2019 beschlossenen Digitalen Versorgungsgesetzes soll durch die Weitergabe von Gesundheitsdaten die Forschung angeregt und betrieben werden (Aponzeller/Orak 2024: 279; Beuth 2022). Das Forschungsdatenzentrum soll hier als eine Zwischenstation fungieren, die die erhaltenen Daten anonymisiert und auf Antrag freigibt (ebd.).

Annähernd zeitgleich stellt Rocher et al. 2019 im Nature Journal vor, dass sie in der Lage sind mithilfe von 15 demografischen Werten 99,98% aller Amerikaner in jedem Datensatz zu identifizieren. Spätestens seit den Enthüllungen von Edward Snowden im Jahre 2013 ist bekannt, dass der Schutz von konkreten Daten oftmals nicht das Problem darstellt, vielmehr besteht das Problem in ihrer ungezügelten Sammlung (Snowden 2019).

Das Ziel dieser Ausarbeitung ist es, das Risiko einer Re-Identifikation von Personen im Kontext von großen Datenbeständen zu bewerten, um die Effektivität von Anonymisierungen zu beurteilen.

2 Grundlagen der Anonymisierung

2.1 Begriffsbestimmung

Bevor Daten im datenschutzkonformen Sinn anonymisiert sind, sind diese zunächst personenbezogenen. Personenbezogene Daten im Sinne von Art. 4 Nummer 1 Datenschutz-Grundverordnung sind alle Informationen, die sich auf eine bereits identifizierte oder identifizierbare Person beziehen. Dabei wird zwischen direkter und indirekter Identifizierbarkeit unterschieden (Schild 2024). Als direkte Merkmale gelten Namen, Anschriften, Telefonnummern, Kennzeichen, Personal- und Sozialversicherungsnummern oder auch eindeutige Berufsbezeichnungen in Kombination mit dem Arbeitgeber (Kinder-Kurlanda/Watteler 2015: 19). Indirekte Merkmale können das Geburtsland, die Staatsangehörigkeit, die Muttersprache, die Berufsbezeichnung oder der Arbeitgeber sein (ebd.).

Damit Organisationen mit personenbezogenen Daten im Rahmen der geltenden Rechtsordnung einfacher arbeiten können, werden diese häufig pseudonymisiert. Pseudonymisierte Daten bleiben personenbezogene Daten. Pseudonyme werden regelmäßig genutzt, um die Wahrscheinlichkeit der Identifikation einer Person zu reduzieren, aber wenn es erforderlich ist, z.B. bei klinischen Studien (Schwartmann et al. 2022: 16), unter der Heranziehung des Pseudonym-Schlüssels umzukehren (Esayas 2015: 8).

An eine Anonymisierung werden stärkere Anforderungen gestellt als an eine Pseudonymisierung (Hamacher et al. 2022: 144). Hier ist eine Bestimmbarkeit unter der Heranziehung weiterer Informationen grundsätzlich nicht möglich. Dabei werden regelmäßig die direkten Merkmale der Identifizierbarkeit gänzlich entfernt (ebd.). Das führt dazu, dass der Anwendungsbereich der Datenschutz-Grundverordnung nicht eröffnet ist, da der Personenbezug neutralisiert wurde (Ulbricht 2015: 186). Das Ziel der Anonymisierung ist demnach, die Möglichkeit eines Rückschlusses auf eine Person zu entfernen (ebd.), um nicht den Anwendungsbereich der Datenschutz-Grundverordnung zu eröffnen.

Die Beurteilung, ob bestimmte Daten rechtlich anonym sind, hängt davon ab, wie unwahrscheinlich eine Re-Identifizierung auf Basis der verfügbaren Technologien, Zeitaufwände und Kosten des jeweiligen Datenbesitzers ist (Europäischer Gerichtshof 2017: 29; Gericht der europäischen Union 2023: 403). Hierbei handelt es sich also stets um einen relativen Maßstab. Dieselben Daten, die in einem Betrieb anonym sind, können in einem anderen personenbezogen sein (Kneuper 2022: 171ff.).

2.2 Die technische Anonymisierung

Neben der rechtlich pragmatischen Abwägung zwischen dem notwendigen Mitteleinsatz und der Wahrscheinlichkeit einer Re-Identifikation, bestimmt sich der Erfolg technischen Anonymisierung negativ: so lange keine Identitäten, z.B. durch verlinkte Datenbestände, offengelegt werden, handelt es sich um anonyme Daten (Hölzel 2018: 502ff.).

Je nachdem, ob Daten in Tabellen oder Teile eines Interviewtranskripts anonymisiert werden sollen, finden sich die Inhalte an verschiedenen Stellen (Kretzer 2013: 3). Während bei Interviewtranskripts die zu anonymisierenden Daten über das gesamte Transkript in unterschiedlichen thematischen Zusammenhängen verteilt sind (ebd.), ist bei Daten in Tabellenform eine spaltenorientierte und/oder zeilenorientierte Perspektive einzunehmen (Goltz et

al. 2017). Sobald die Daten identifiziert worden sind, können diese durch verschiedene Anonymisierungstechniken anonymisiert werden (Krebs/Hagenweiler 2022: 82).

Davor ist zu klären mit welchem Fokus die Anonymisierung durchgeführt werden soll: eine Anonymisierung steht stets im Spannungsverhältnis zwischen Informationsgehalt und Robustheit der Anonymisierung (Wilhelm et al. 2023: 37).

Abbildung 1: Robustheit/Informationsgehalt

Tabelle 1: Fiktiver Beispieldatensatz mit Fokus auf Robustheit anonymisiert

Name	Alter	Geschl.	Arbeitsbereich	PLZ	Std./Woche	Einkommen
***	20-29	m	Dienstl. und Handwerk	5****	31-40	30.000 - 39.999
***	30-39	m	Dienstl. und Handwerk	4****	>40	> 50.000
***	50-59	w	Büro und Verwaltung	2****	31-40	> 50.000
***	30-39	w	Büro und Verwaltung	5****	31-40	30.000 - 39.999
***	30-39	w	Kunst und Kultur	4****	21-30	< 19.999

Tabelle 2: Fiktiver Beispieldatensatz mit Fokus auf Informationsgehalt anonymisiert

Name	Alter	Geschl.	Arbeitsbereich	PLZ	Std./Woche	Einkommen
W. E****	24	m	Postbote	56284	39	34.100
F. W******	30	m	sebstst. Handwerker	42113	50	64.300
J. S*****	56	w	Steuerberaterin	24626	40	79.540
S. B***	38	w	Bürokauffrau	54450	39	38.500
K. H***	33	w	Musikerin	43221	20	19.800

Quelle: Wilhelm et al. 2023: 37.

Die elementaren Anonymisierungstechniken sind die Randomisierung und die Generalisierung (Krebs/Hagenweiler 2022: 83-86). Im Rahmen der Randomisierung werden Daten innerhalb eines Datensatzes so verfälscht, dass eine Verbindung zwischen Daten und betroffener Person nicht mehr möglich ist. Dazu werden die Daten entweder stochastisch überlagert, verfälscht, gelöscht oder geclustert (ebd.). Bei der stochastischen Überlagerung werden entweder zufällig oder absichtlich einzelne Daten so verändert, dass sie, bei Erhaltung der allgemeinen Verteilung im Gesamtdatenbestand, weniger genau sind (Art.-29-Datenschutzgruppe 2014: 14f.). Bei der Verfälschung werden dem Datenbestand künstliche Werte hinzugefügt oder die einzelnen Werte statistisch geändert. Diese Methode führt zu einem starken Nutzbarkeitsverlust der Daten (ebd.: 16). Bei der Löschung werden atypische Daten im Bestand entfernt, die sonst zu einer direkten Identifikation einer Person geführt hätten (ebd.:

35). Bei dem Clustering werden Datensätze in Unterklassen gruppiert, wobei die Verteilung der Attributwerte weiterhin repräsentativ bleib (Hölzel 2018: 505).

Gerichtsurteile werden regelmäßig mittels Randomisierung anonymisiert (Deuber/Keuchen 2023).

Bei der Generalisierung hingegen werden die Daten durch eine Veränderung des Größenintervalls durch einen weniger spezifischen Wert ersetzt (Krebs/Hagenweiler 2022: 86f.; Art.-29-Datenschutzgruppe 2014: 19, 39ff.).

3 Re-Identifikation

3.1 Begriffsbestimmung

Sobald es Akteuren, absichtlich oder zufällig, gelingt, einen Personenbezug in einem bislang für anonym gehaltenen Datenbestand herzustellen, wird von einer De-Anonymisierung bzw. Re-Identifikation der anonymen Daten gesprochen (Sorge 2013; Vokinger/Muehlematter 2019; Appenzeller/Orak 2024). Hierfür gibt es anfragebasierte und datenbasierte Verfahren (Goltz et al. 2017).

Bei anfragebasierten Verfahren führt der Angreifer eine Reihe bestimmter Anfragen innerhalb einer Datenbank aus, um versteckte Relationen aufzudecken. Datenbankmanagementsysteme bieten die Möglichkeit, dass Nutzende Anfragen stellen können, um eine dedizierte Sicht auf die Datensätze zu erhalten. Diese Funktion wird dann zweckentfremdet (ebd.).

Bei einem datenbasierten Verfahren wird der Fokus auf die Daten gesetzt (ebd.). Ziel ist es, die technische Anonymisierung umzukehren. Dazu wird versucht herauszufinden, wie hoch der Aggregationsgrad der Daten ausgeprägt ist (für aggregationsbasierte Verfahren vgl. Krebs/Hagenweiler 2022: 87-93; k-anonymity, l-Diversity, t-closeness, δ-Presence). Wenn klar ist, wie stark die Daten aggregiert worden sind oder mit welchen Techniken die Daten anonymisiert worden sind, wird ein Re-Identifikationsversuch durchgeführt (Universität Washington 2019: 3). Vermehrt werden auch sogenannte Linkage-Attacken durchgeführt mit der die anonymisierten Daten mit weiteren zu Verfügung stehenden Daten korreliert werden (Hamacher et al. 2022: 145; Hölzel 2018: 502ff.).

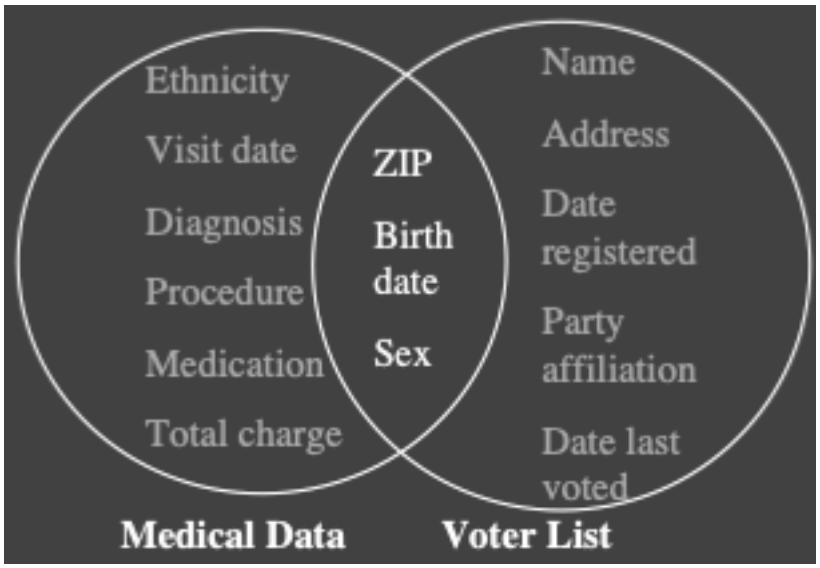
Die Motivation und der Typ des Angreifers können dabei variieren: ein Wissenschaftler könnte dabei die Motivation haben, eine Re-Identifikation herbeizuführen, um eine formulierte Hypothese zu testen, die mit den zu Verfügung stehenden Daten nicht falsifizierbar

wäre (ebd.). Datenanalysten könnten herausfinden wollen, ob der zu Grunde liegende Datenbestand tatsächlich anonym ist (ebd.), denn die Anonymisierungsqualität lässt sich rechtlich und technisch nur empirisch ermitteln (Deuber/Keuchen 2023). Kriminelle könnten ein Interesse an den Daten haben, um einen Identitätsdiebstahl zu begehen und wirtschaftlich handelnde Akteure könnten ein Interesse daran haben, personenbezogene Daten zu sammeln, um diese zu monetarisieren (Universität Washington 2019: 3).

3.2 Beispiele von Re-Identifikationen

Das frühste Beispiel einer prominenten Re-Identifikation fand im Jahr 2002 in Massachusetts statt (Sweeney 2002: 557ff.). Die Group Insurance Commission war seinerzeit für die Beschaffung der Krankenversicherung der 135.000 Angestellten und Familienangehörigen des Staates verantwortlich. Da die Organisation die Daten zuvor anonymisiert hatte, verkaufte sie die Daten an die Industrie und gab ebenfalls ein Exemplar zur Forschung frei. Latanya Sweeney bewies 2002, dass Sie in der Lage war, mit einem Auszug des Wahlregisters von Massachusetts aus dem Jahre 1997 diese Daten mit den Daten der Group Insurance Commision zu verlinken und Identitäten über die Postleitzahl, das Geschlecht und das Geburtsdatum zu re-identifizieren. Das ermöglichte Sweeney darzustellen, welche Diagnosen, Medikamente oder Prozeduren innerhalb des Gesundheitsdatenbestandes vorkamen und zu welchem Individuum die Daten gehörten. Eine der identifizierten Personen war der damalige Gouverneur von Massachusetts William Weld (ebd.).

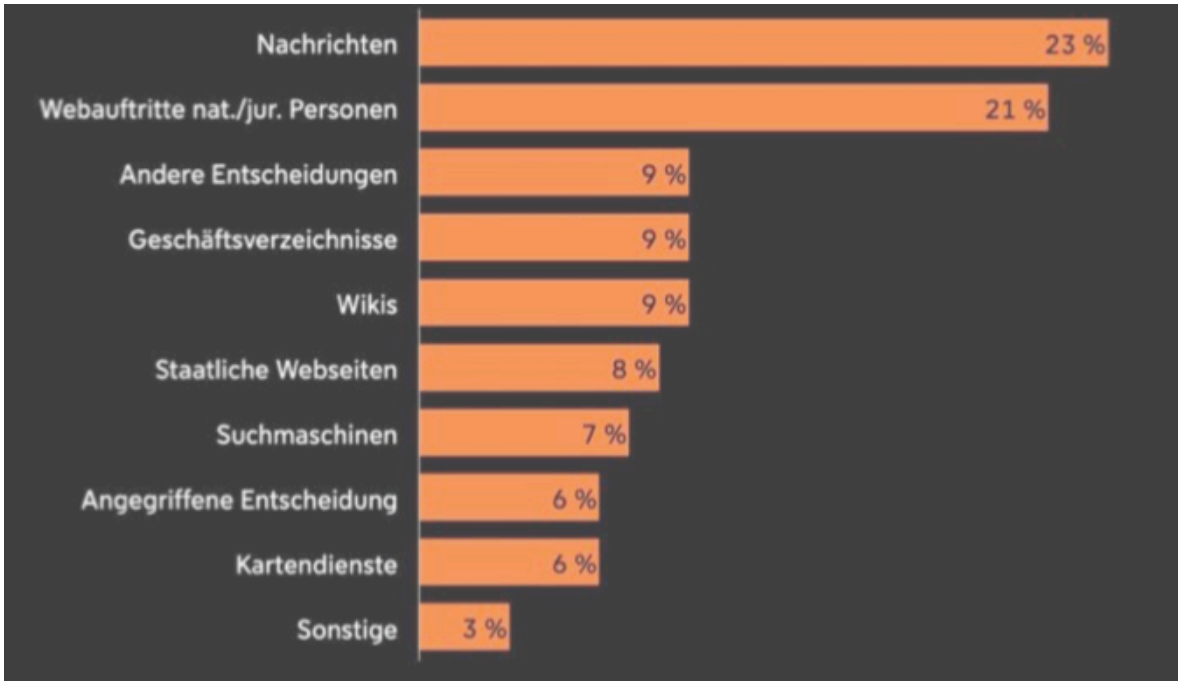
Abbildung 2: Data-Linkage



Quelle: Sweeney 2002: 559.

Ein weiteres prominentes Beispiel ereignete sich im Jahr 2006: der damalige DVD-Verleihservice Netflix einen Wettbewerb mit einem Preisgeld von einer Millionen Dollar veranstaltete, mit dem Zweck, die Filmbewertungsmechanismen zu verbessern (Hafner 2006). Zur Unterstützung veröffentliche Netflix 100 Millionen Filmbewertungen die von 480.000 Nutzern während der Jahre 1999 – 2005 gemacht wurden (Narayanan/Shmatikov 2007). Dabei wurden alle identifizierenden Daten entfernt. Der Datensatz bestand aus einer Filmbewertung und einem dazugehörigen Datum. Narayan und Shmatikov stellten dar, dass nur sehr wenig Hilfs-Informationen notwendig sind, um die Daten zu re-identifizieren. Dafür konnten Meldungen auf persönlichen Internetblogs oder die gemachten Filmkritiken bei anderen Internet-Filmdatenbanken wie IMDb genutzt werden. Dabei wurde zu Grunde gelegt, dass Nutzer, die Kritiken auf Netflix veröffentlichten, meist auch auf anderen Kritikportalen angemeldet waren und dass diese öffentlichen Kritiken sachlich und zeitlich nah beieinander liegen müssten. Dabei zeigte sich, dass zwei Korrelationen von Bewertungen und Datumsangaben mit einer Fehlerquote von +/- drei Tagen ausreichend waren, um 68% der Nutzer im Netflix-Datensatz zu identifizieren. Mit acht Bewertungen und Datumsangaben inklusive einer 14-tägigen Fehlerquote waren sogar 99% der Nutzer identifizierbar (ebd.).

Abbildung 3: unterschiedliche Quellen



Quelle: Deuber/Keuchen 2023.

Besonders interessant sind auch die Beispiele aus dem Bereich der re-identifizierten Gerichtsentscheidungen (Vokinger/Mühlematter 2019; Deuber/Keuchen 2023). Im Beispiel

von Deuber und Keuchen wurden 54 Jurastudenten drei Stunden lang, ohne spezifische Anleitung oder Training und nur mit Hilfe des Internets, angeheuert um Gerichtsentscheidungen zu re-identifizieren. Dazu wurden 50 Gerichtsentscheidungen mit insgesamt 484 einzigartigen anonymisierten Informationen zu natürlichen und juristischen Personen präsentiert. Für die Re-Identifikation nutzten die Studierenden eine Vielzahl von unterschiedlichen Quellen (vgl. Abb. 3).

In 115 von den 484 anonymisierten Informationen konnte das Ergebnis der potenziellen Re-Identifikation des Studierenden vom Forschungsleiter bestätigt werden. Dabei war mehr als jede dritte Information der Klarname einer natürlichen oder juristischen Person (ebd.).

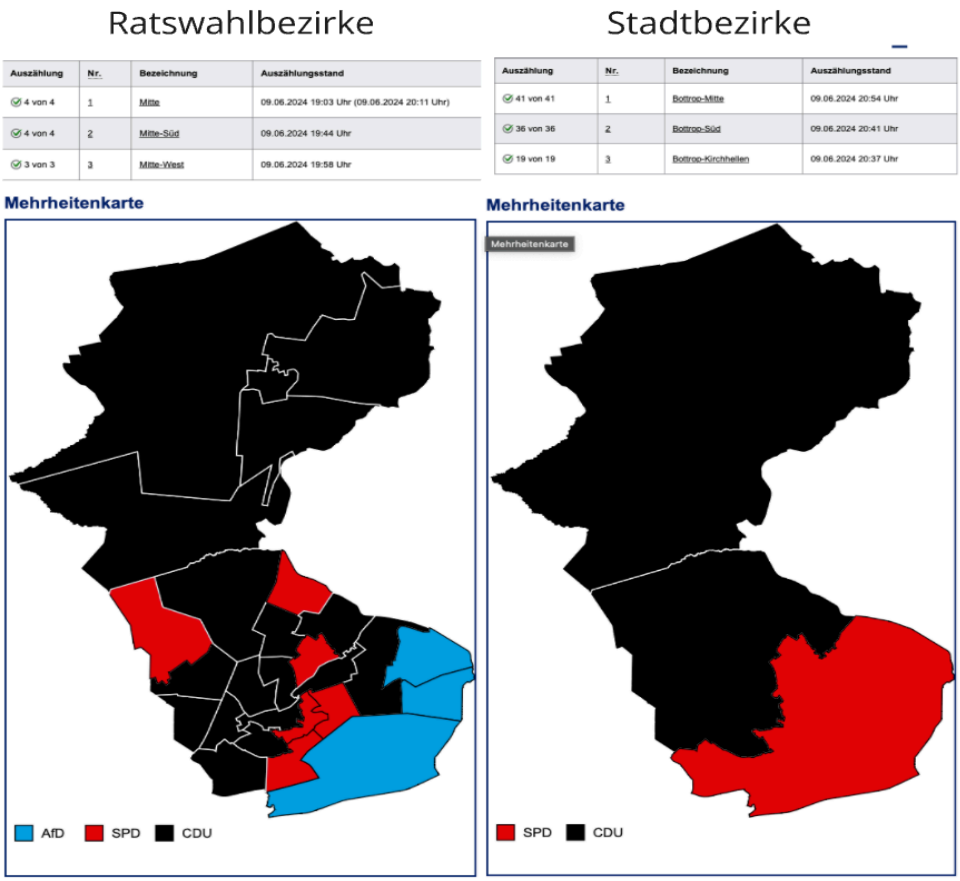
Auch der schweizerische Fall aus dem Jahre 2019 bei dem Vokinger und Muehlematter bewiesen haben, dass auch ganz konkrete Fragestellungen, die sich hinter vielen anonymisierten Gerichtsentscheidungen verbergen, zuverlässig beantwortet werden können: in dem vorliegenden Fall wurden 122.218 Entscheidungen des Bundesgerichts und 58.732 Entscheidungen des Bundesverwaltungsgericht von 2000 – 2018 nach den Verfahren durchsucht, bei denen pharmazeutische Unternehmen wegen ihrer Arzneimittel gegen eine Preisverfügung des Bundesamts für Gesundheit geklagt hatten. Daraufhin wurden diese Verfahren, inkl. der verfügbaren erstinstanzlichen Urteile, auf direkte oder indirekte Merkmale untersucht. In 21 von 25 untersuchten Fällen wurden die pharmazeutischen Unternehmen mit dem jeweiligen Arzneimittel identifiziert. In jedem der Beispiele fand die Re-Identifikation über verlinkte Datenbestände statt.

3.3 Praxisversuch Re-Identifikation

Im nun folgenden Teil wird ein Praxistest zur Realisierbarkeit von Re-Identifikationen über verlinkte Datenbestände durchgeführt. Dazu möchte der Verfasser der Frage nachgehen, wie und ob es möglich ist, eine Re-Identifikation eines veröffentlichten Wahlergebnisses durchzuführen. Als Praxisbeispiel dient hierzu das Wahlergebnis der Europawahl 2024 für Bottrop. Die Wahldaten zur Europawahl werden auf der städtischen Webseite aggregiert publiziert und die Sieger werden geografisch in Stadtbezirke und Ratswahlbezirke dargestellt. Zusätzlich werden die Ergebnisse der einzelnen Wahlbezirke (auch Stimmbezirk genannt) als Liste analog zu den Auszügen der Listen in Abbildung 4 veröffentlicht.

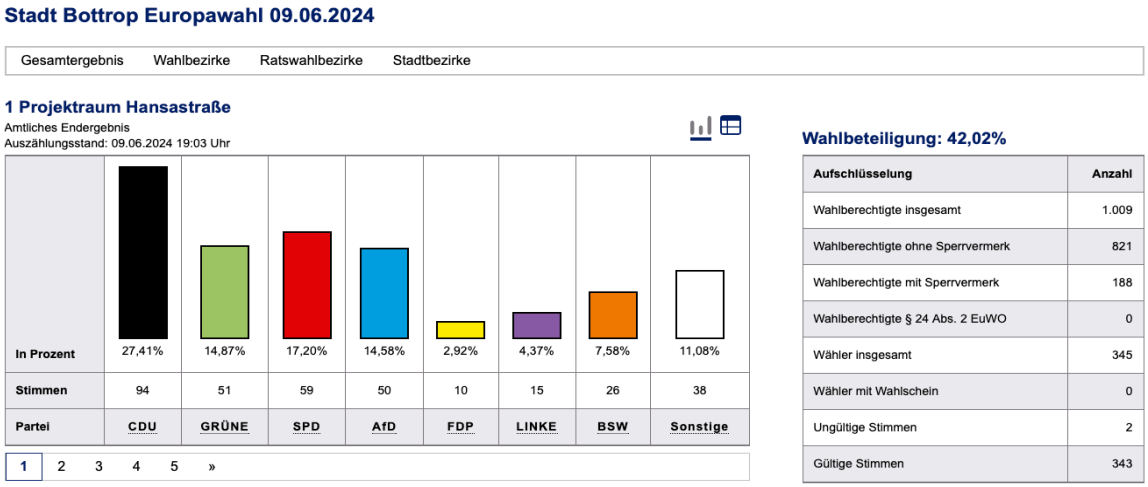
Bei Betrachtung der Wahlergebnisse von Wahlbezirk, Ratswahlbezirk und Stadtbezirk zeigt sich, dass nur wenig direkte oder indirekte Merkmale ableitbar sind, die zur Identifikation einzelner Subjekte dienen könnten.

Abbildung 4: Wahlergebnispräsentation



Quelle: Stadt Bottrop 2024a, 2024b

Abbildung 5: Wahlbezirk-Hansastraße



Quelle: Stadt Bottrop 2024c

Grundsätzlich können nutzbare Informationen aus der Eigenschaft der Wahlberechtigung abgeleitet werden, denn um bei der Europawahl wahlberechtigt zu sein, muss man Deutscher im Sinne des Grundgesetztes sein oder Staatsangehöriger eines EU-Mitgliedsstaats sowie am Wahltag das 16 Lebensjahr vollendet haben (Bundeswahlleiterin 2024). D.h. im Einzugsgebiet des Wahlbezirkes aus Abbildung 5 leben 1009 Deutsche bzw. EU-Ausländer, die das 16. Lebensjahr vollendet haben.

Um weitere Analysen anstellen zu können, benötigt es weitere demografische Informationen über die Gruppe der Wahlberechtigten. Diese lassen sich über das Statistikamt der Stadt Bottrop (2024d) beziehen.

Aus der Bevölkerungsstatistik (ebd.). ergehen vier wesentliche Informationen über geografisch abgegrenzte statistische Bezirke: a) die Zahl der Wohnbevölkerung mit dem Lebensmittelpunkt in Bottrop, b) eine Unterscheidung in Männern und Frauen c) die Zugehörigkeit zum statistischen Bezirk (geografisch abgegrenzte Einheit) und d) die Unterscheidung in Deutsche und Ausländer.

Abbildung 6: Bevölkerungsstatistik-2023

Stadt Bottrop
Amt für Informationsverarbeitung (12)

Bevölkerungsstatistik

Stichtag: 31.12.202

Statistischer Bezirk		Deutsche			Ausländer			Wohnbevölkerung gesamt		
Nr.	Bezeichnung	männl.*	weibl.	gesamt	männl.*	weibl.	gesamt	männl.*	weibl.	gesamt
11	Altstadt	1 894	2 053	3 947	878	838	1 716	2 772	2 891	5 663
12	Nord-Ost	3 683	3 702	7 385	1 061	1 039	2 100	4 744	4 741	9 485
13	Süd-West	4 566	5 092	9 658	833	864	1 697	5 399	5 956	11 355
21	Fuhlenbrock-Heide	2 028	2 228	4 256	135	142	277	2 163	2 370	4 533
22	Fuhlenbrock-Wald	4 003	4 401	8 404	241	280	521	4 244	4 681	8 925
31	Stadtwald	1 638	1 779	3 417	93	68	161	1 731	1 847	3 578
32	Eigen	5 266	5 685	10 951	757	621	1 378	6 023	6 306	12 329
41	Batenbrock-Nord	4 022	4 132	8 154	658	587	1 245	4 680	4 719	9 399
42	Batenbrock-Süd	3 794	4 126	7 920	1 118	1 085	2 203	4 912	5 211	10 123
51	Boy	3 787	3 913	7 700	783	680	1 463	4 570	4 593	9 163
52	Welheim	1 720	1 797	3 517	533	522	1 055	2 253	2 319	4 572
61	Ebel / Welheimer Mark	1 204	1 176	2 380	197	226	423	1 401	1 402	2 803
62	Süd	2 263	2 398	4 661	305	305	610	2 568	2 703	5 271
Alt-Bottrop gesamt		**39 868**	**42 482**	**82 350**	**7 592**	**7 257**	**14 849**	**47 460**	**49 739**	**97 199**
71	Kirchhellen-Mitte	5 157	5 686	10 843	195	226	421	5 352	5 912	11 264
72	Kirchhellen-Süd / Grafenwald	2 657	2 811	5 468	144	127	271	2 801	2 938	5 739
73	Kirchhellen-Nord-West	676	676	1 352	79	45	124	755	721	1 476
74	Kirchhellen-Nord-Ost	1 257	1 255	2 512	41	44	85	1 298	1 299	2 597
Kirchhellen gesamt		**9 747**	**10 428**	**20 175**	**459**	**442**	**901**	**10 206**	**10 870**	**21 076**
Stadt Bottrop gesamt		**49 615**	**52 910**	**102 525**	**8 051**	**7 699**	**15 750**	**57 666**	**60 609**	**118 275**
zum Vergleich 31.12.2022 gesamt:		**49 696**	**53 158**	**102 854**	**7 465**	**7 309**	**14 774**	**57 161**	**60 467**	**117 628**

* Personen mit Registereintrag "divers" wurden aus Gründen des Statistikgeheimnisses zur männlichen Bevölkerung gezählt.

Quelle: Stadt Bottrop 2024d

Durch eine Recherche über den statistischen Bezirk ist dem Verfasser aufgefallen, dass die Stadt Bottrop im sog. *Stadtprofil* (Stadt Bottrop 2024e: 18f.) auch eine dedizierte Altersstruktur über die statistischen Bezirke präsentiert (vgl. Abb.7).

Die Arbeitshypothese lautet: sofern es möglich ist, den statistischen Bezirk mit einem Ratsbezirk, Wahlbezirk oder Stadtbezirk zu verlinken, würde man demografische Informationen über die Zusammensetzung der Wahlberechtigten erhalten. Denn die Wahlberechtigung einer Person tritt qua Gesetz ein. Hierzu ist keine Handlung der jeweiligen Person erforderlich und wie aus Abbildung 5 zu erkennen ist, werden alle Personen, die wahlberechtigt sind, gelistet. Dazu ist im nächsten Schritt der geografische Zuschnitt der statistischen Bezirke mit einem Wahlbezirk zu korrelieren.

Der geographische Zuschnitt der statistischen Bezirke, sowie der Zuschnitt der Wahlbezirke ist über das Open-Data-Portal der Stadt Bottrop zu beziehen (Stadt Bottrop 2024f; 2024g). Diese lassen sich durch ein Geoinformationssystem (QGIS) digital visualisieren. Dazu werden beide Datensätze in die Open-Source-Software QGIS eingeladen und miteinander vergleichen, um herauszufinden, ob es Bezirke gibt, die annähernd gleich geschnitten sind.

Abbildung 7: Stadtprofil

Merkmale	Statistischer Bezirk 21 Fuhlenbrock-Heide
Fläche und Bevölkerung	
Fläche in qkm	1,22
Bevölkerungsdichte (Einw. je qkm)	3 716
Einwohner insgesamt	4 533
männlich	2 163
weiblich	2 370
Deutsche	4 256
Einw. mit Migrationshintergrund	741
Migrantenanteil in %	16,3
darunter Ausländer	277
Ausländeranteil in %	6,1

Merkmale	Statistischer Bezirk 21 Fuhlenbrock-Heide
Altersaufbau der Bevölkerung	
0 bis unter 3 Jahre	101
in %	2,2
3 bis unter 6 Jahre	120
in %	2,6
6 bis unter 10 Jahre	139
in %	3,1
10 bis unter 18 Jahre	251
in %	5,5
18 bis unter 21 Jahre	103
in %	2,3

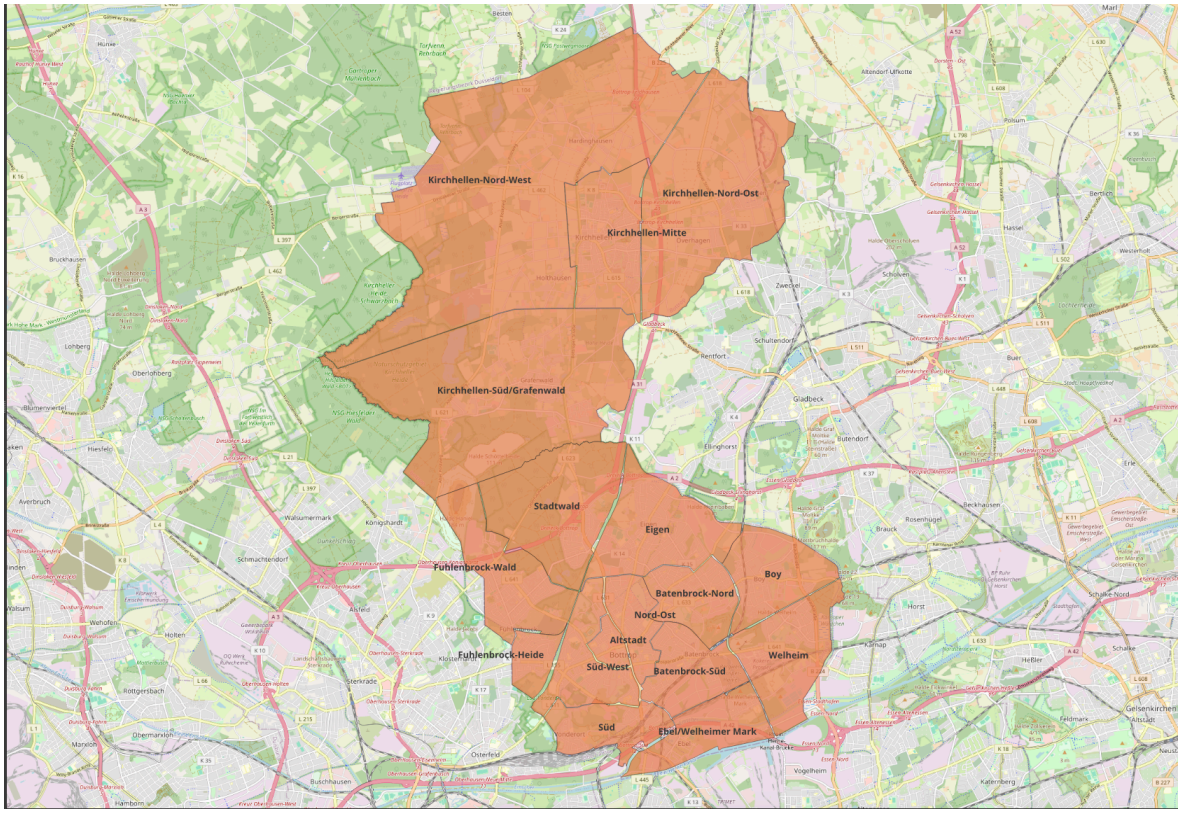
Merkmale	Statistischer Bezirk 21 Fuhlenbrock-Heide
21 bis unter 30 Jahre	354
in %	7,8
30 bis unter 45 Jahre	867
in %	19,1
45 bis unter 65 Jahre	1 283
in %	28,3
65 bis unter 80 Jahre	900
in %	19,9
80 Jahre und älter	415
in %	9,2

Quelle: Stadt Bottrop 2024e

Der gleiche Prozess aus Abbildung 8 wird nun mit dem Zuschnitt der Ratswahlbezirke in Abbildung 9 durchgeführt. Beide Zuschnitte lassen sich übereinanderlegen, um die Bezirke zu identifizieren, die einen gleichen geographischen Zuschnitt haben (vgl. Abb. 10). Zur besseren Lesbarkeit wurden die Visualisierungseigenschaften etwas angepasst.

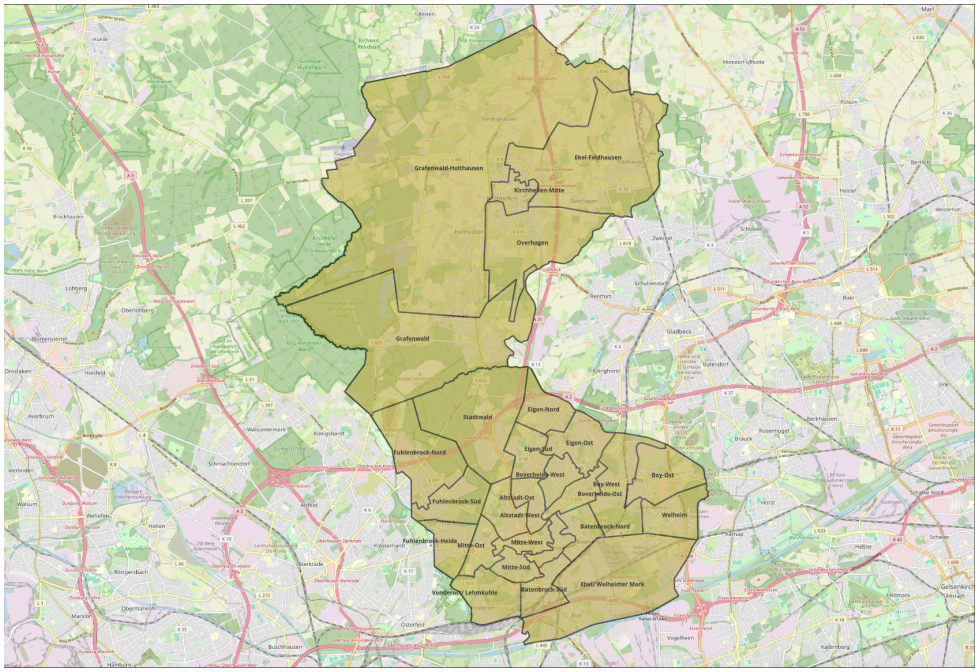
Insbesondere der geographische Zuschnitt des statistischen Bezirks Fuhlenbrock-Heide (unten links in Abb. 10) und der des gleichnamigen Ratswahlbezirk überschneiden sich nahezu identisch.

Abbildung 8: statistische Bezirke



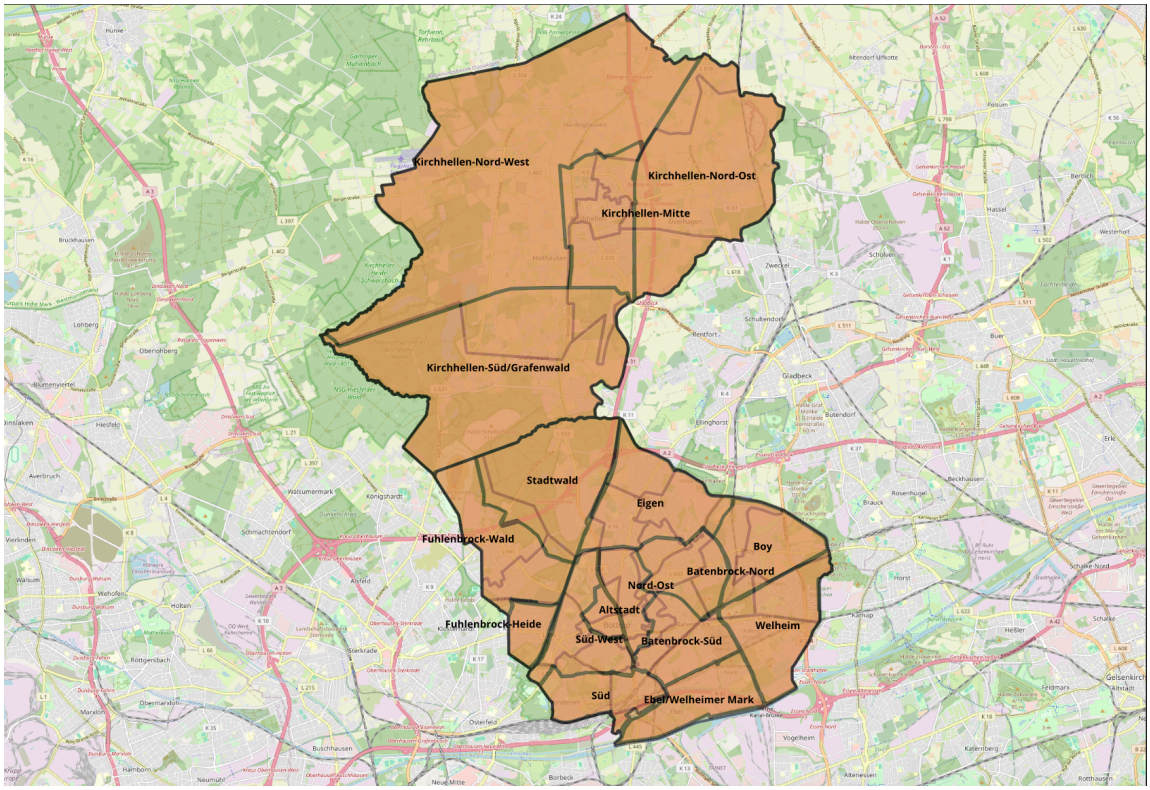
Quelle: eigene Abbildung mit Stadt Bottrop 2024f

Abbildung 9: Ratswahlbezirke



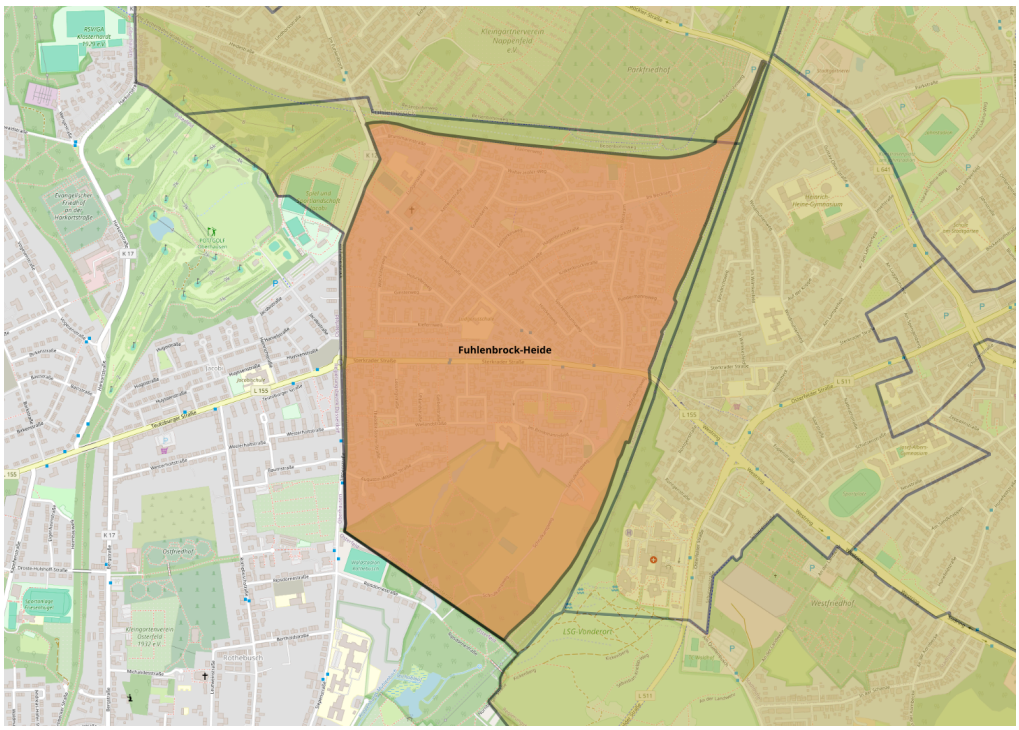
Quelle: eigene Abbildung mit Stadt Bottrop 2024g

Abbildung 10: Bezirke überlagert



Quelle: eigene Abbildung mit Stadt Bottrop 2024f; 2024g

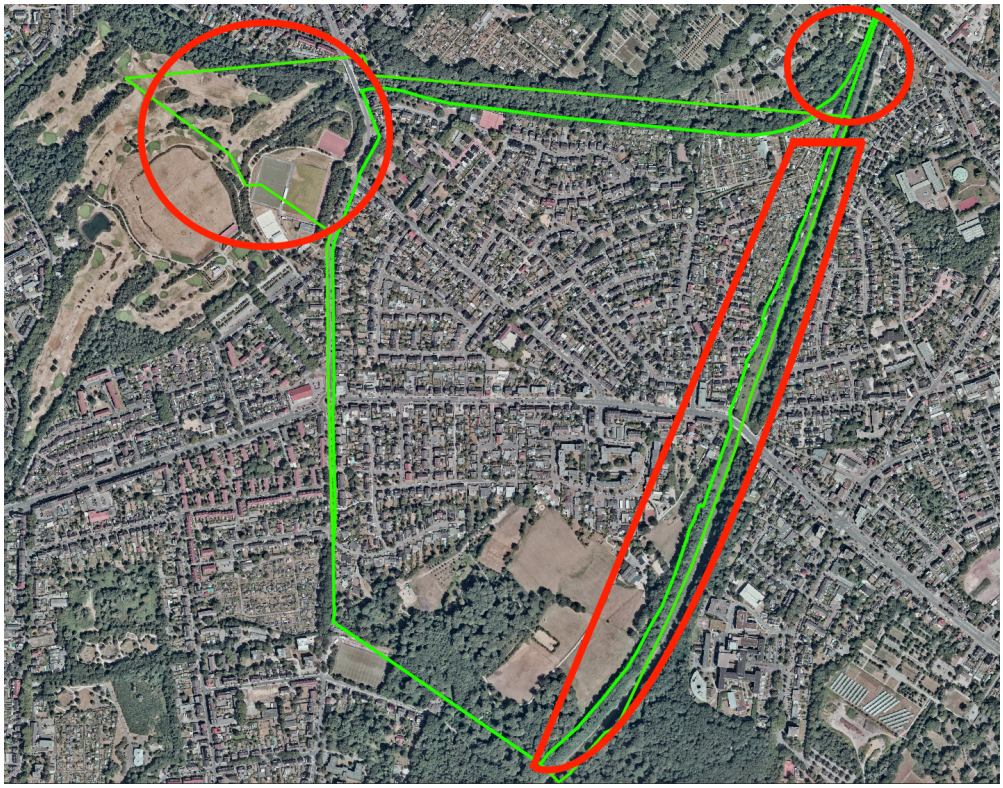
Abbildung 11: Fuhlenbrock-Heide



Quelle: eigene Abbildung mit Stadt Bottrop 2024f; 2024g

Auch unter Zuhilfenahme des nordrheinwestfälischen Luftbildes (Geobasis NRW 2024) zeigt sich, dass sich in den Grenzbereichen (rotgefärbt in Abbildung 12), bei denen der Ratswahlbezirk größer als der statistische Bezirk ist, keine Wohnhäuser sind. Demnach handelt es sich im Kontext der dort lebenden Personen um dieselbe Anzahl und somit auch um dieselbe Anzahl an Wahlberechtigten (vgl. Abbildung 13).

Abbildung 12: Luftbild

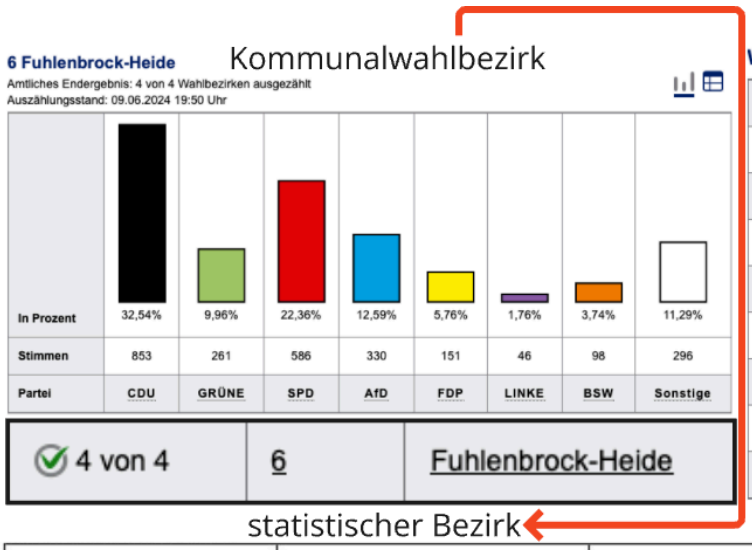


Quelle: eigene Abbildung mit Stadt Bottrop 2024f; 2024g; Geobasis NRW 2024

Durch die Korrelation zwischen den Bezirken (vgl. Abb. 13) können die demografischen Informationen des Stadtprofils für den Bezirk Fuhlenbrock-Heide aus Abbildung 7/11 als gegeben betrachtet werden. Um weiter mit der Re-Identifikation voranzuschreiten, wird sich für den Praxisversuch auf die Re-Identifikation der Wähler der Linkspartei eingeschlossen.

Die Linkspartei hat einen kleinen Stimmenteil erhalten, der aus Sicht des Verfassers das größte Potential für eine Umkehr der Anonymisierung bietet, sofern weitere geeignete Daten vorliegen. Darüber hinaus wird es im Kontext dieser Abhandlung als notwendig betrachtet, das untersuchte Thema weiter einzugrenzen. Dabei stellt sich der Verfasser die Frage, wie viele junge Menschen im Alter von 16 bis unter 30 die Linkspartei im Bezirk Fuhlenbrock-Heide gewählt haben.

Abbildung 13: Korrelationen



Wahlbeteiligung: 70,10%

Aufschlüsselung	Anzahl
Wahlberechtigte insgesamt	3.763
Wahlberechtigte ohne Sperrvermerk	2.757
Wahlberechtigte mit Sperrvermerk	1.006
Wahlberechtigte § 24 Abs. 2 EuWO	0
Wähler insgesamt	2.638
Wähler mit Wahlschein	933
Ungültige Stimmen	17
Gültige Stimmen	2.621

Statistischer Bezirk		Deutsche			Ausländer			Wohnbevölkerung gesamt		
Nr.	Bezeichnung	männl.*	weibl.	gesamt	männl.*	weibl.	gesamt	männl.*	weibl.	gesamt
21	Fuhlenbrock-Heide	2 028	2 228	4 256	135	142	277	2 163	2 370	4 533

Quelle: eigene Abbildung mit Stadt Bottrop 2024d; 2024h

Wie aus Abbildung 6, 7 und 14 hervorgeht, liegt die Zahl der Wohnbevölkerung bei 4533 Personen. Hierbei eingeschlossen sind alle Personen, die in der Stadt wohnen.

Im ersten Schritt ist daher zunächst die Zahl der wahlberechtigten Personen zu ermitteln. Der erste Indikator dafür ist die Altersverteilung in der Wohnbevölkerung. Nur Personen, die das 16. Lebensjahr vollendet haben, sind grundsätzlich im wahlberechtigten Alter.

Damit fallen nach der demografischen Verteilung ohne jegliche Berechnung bereits 360 Personen heraus (0-3 Jahre=101, 3-6 Jahre=120, 6-10=139 – vgl. Abb. 14).

Um den Anteil der Personen von 10-15 Jahren zu ermitteln, wird die Anzahl des Intervall *10 bis unter 18 Jahre* durch acht geteilt und mit sechs multipliziert, um dem Durchschnitt nach die 16- unter 18-Jährigen vom Intervall zu trennen. Demnach liegt die Anzahl der Wohnbevölkerung, die mindestens 16 Jahre alt sind, bei 3985 (548 Personen fallen weg; davon 188 unter 16-Jährige; somit 63 Heranwachsende in der gesuchten Spanne).

Bei Beibehaltung der prozentualen Verteilung zwischen Deutschen und Ausländern aus Abbildung 6/13 handelt es sich dabei um 3742 Deutsche (93,9%) und 243 (6,1%) Ausländer. Dabei sind die Deutschen Personen grundsätzlich wahlberechtigt. Bei den Ausländern ist noch eine Bereinigung nach EU-Ausländer und Nicht-EU-Ausländer durchzuführen, um die

Wahlberechtigten Ausländer zu identifizieren. Dazu wird die Differenz aus den wahlberechtigten des Bezirks (vgl. Abb. 13) und den 3742 Deutschen gebildet. Hiernach ergibt sich eine Anzahl i.H.v. 19 (7,82% der Ausländer) wahlberechtigten EU-Ausländern.

Abbildung 14: Altersaufbau

Altersaufbau der Bevölkerung	
0 bis unter 3 Jahre	101
in %	2,2
3 bis unter 6 Jahre	120
in %	2,6
6 bis unter 10 Jahre	139
in %	3,1
10 bis unter 18 Jahre	251
in %	5,5
18 bis unter 21 Jahre	103
in %	2,3
21 bis unter 30 Jahre	354
in %	7,8
30 bis unter 45 Jahre	867
in %	19,1
45 bis unter 65 Jahre	1 283
in %	28,3
65 bis unter 80 Jahre	900
in %	19,9
80 Jahre und älter	415
in %	9,2

Quelle: Stadt Bottrop 2024e

Daraus lässt sich ermitteln, dass 0,51% der wahlberechtigten Personen EU-Ausländer im Bezirk sind. Da auf die Linken im Bezirk nur 46 Stimmen entfallen sind (vgl. Abb. 13) hat mit sehr hoher Wahrscheinlichkeit kein wahlberechtigter EU-Ausländer im Bezirk die Linken gewählt.

Die Wohnbevölkerung der 16 – unter 30-jährigen im Bezirk besteht aus 520 Personen (103+354+63). Hierbei handelt es sich somit um 488 (93,9%) deutsche wahlberechtigte Personen. Von den übrigen 32 ausländischen Personen sind drei (2,5024; 7,82%; vgl. oben) EU-Ausländer. Die Gesamtzahl der wahlberechtigten 16- unter 30-jährigen Personen beträgt demnach 491. Da die Wahlbeteiligung bei 70,1% lag, liegt die Zahl der abgegebenen Stimmen bei 344, wovon rechnerisch 2 Stimmen (0,64%) ungültig waren (vgl. Abb. 13). Damit hat die betrachtete Gruppe im Bezirk 342 gültige Stimmen abgegeben wovon 1,76% auf die

Linkspartei entfallen sind. Hiernach wurde in der betrachten Gruppe reinrechnerisch sechsmal die Linkspartei gewählt. Dabei handelt es sich reinrechnerisch um 3,17 Frauen und 2,83 Männer mit deutscher Staatsangehörigkeit (vgl. Abb. 6).

Abbildung 15: möglicher Treffer

Quelle: Stadt Bottrop 2024i

Durch die Liste der Wahlvorschläge für die Wahl in den Kommunalwahlbezirken (Stadt Bottrop 2024i) und einer gezielten Suche über die Postleitzahl des beobachteten Bezirks, ist es dem Verfasser gelungen, wohlmöglich eine natürliche Person zu identifizieren, die eine von sechs jungen Linkswählern sein könnte. Dieser Umstand müsste weiter geprüft und verifiziert werden, indem weitere Datenbestände herangezogen werden, um das Ergebnis zu valideren.

4 Fazit

Zusammenfassend lässt sich sagen, dass eine vollständige Anonymität praktisch nur dann gewährleistet werden kann, solange die anonymisierten Daten den Machtbereich des Datenhalters noch nicht verlassen haben (Hamacher et al. 2022: 145). Durch die Heranziehung weiterer Datenquellen lässt sich das Risiko einer Re-Identifikation nie ausschließen. Dabei darf dieser Umstand nicht dazu führen, dass das Informationsbedürfnis der Öffentlichkeit auf Grund einzelner Bedenken zurücktritt (ebd.).

Daher führt dieser Umstand dazu, dass eine Beschränkung des Datenschutzes auf ausschließlich personenbezogene Daten nicht mehr als zeitgemäß erscheint (Kneuper 2022: 171). Darum sollten auch anonymisierte Daten als schützenswert definiert werden. Diese Forderung könnte mit einem gesetzlichen Verbot einer Re-Identifikation gepaart werden (ebd.; auch Vokinger/Muehlematter 2019).

Aber es gibt auch weitere Möglichkeiten den Schutz von Betroffenen künftig zu regeln: so wird bereits jetzt bei den anonymen Daten des Mikrozensus eine vollständige Datenansicht nur an Arbeitsplätzen vor Ort angeboten (Watteler/Kinder-Kurlanda 2015: 516f.). Für eine Datennutzung außerhalb der Ämter werden nur stark verkürzte Stichproben herausgegeben (ebd.).

Eine weitere Idee wurde vom Erfinder des World Wide Web Sir Tim Berners Lee vorgetragen (Robinson 2022). Die Solid Technologie verfolgt das Ziel, dass alle gesammelten persönlichen Daten einer jeden natürlichen Personen in höchstpersönlichen Datentöpfen gespeichert werden. Jede natürliche Person ist selbst in der Lage auf Anfrage Zugriff zu ihren Datentöpfen zu gewähren und diesen jederzeit zu entziehen (ebd.). Die hergestellte Datenhoheit könnte zu mehr Sensibilität führen.

Im Kontext des Praxisversuchs hingegen wurde gezeigt, dass bereits ohne großes technisches Knowhow die Wahrscheinlichkeit für eine potenzielle Re-Identifikation ansteigt, sobald genügend verknüpfbare Daten vorliegen. Dabei ist nicht auszuschließen, dass das Experiment im Kontext der Wahrscheinlichkeitsrechnungen Fehlannahmen unterlag, diese dürften aber im Kontext eines nicht akribisch geplanten Re-Identifikationsversuchs zu vernachlässigen sein. In Nachfolgeexperimenten könnte dieser Umstand einen wertvollen Ansatzpunkt darstellen, um die Re-Identifikation zu validieren. Durch die Vernetzung des statistischen Bezirks mit dem Ratswahlbezirk werden viele demografische Daten der Wählenden über ihr Wahlverhalten im beobachteten Raum freigestellt.

Hierbei erschien es günstig, dass der statistische Bezirk einen annähernd gleichen geografischen Zuschnitt wie der Ratswahlbezirk hatte. Dennoch erscheint es aber auch nicht als zwingende Voraussetzung, dass beide Bezirke miteinander deckungsgleich sind, um zuverlässige Annäherungen oder Faktoren für eine potenzielle Re-Identifikation abzuleiten. Auch wurden hierbei keinerlei Informationen aus den sozialen Medien berücksichtigt noch besondere Software oder Künstliche Intelligenz genutzt, die beim Errechnen von Wahrscheinlichkeiten oder dem Vernetzten von Datenpunkten behilflich sein könnte. Während der Bearbeitung hat sich gezeigt, dass Standortdaten im Allgemeinen überdurchschnittlich wertvoll bei Re-Identifikationen sein können: sobald z.B. gesichert ist, dass eine Person am Wahltag tatsächlich in einem Stimmbezirksbüro (oder in dessen Nähe) gewesen ist, erhöht sich die Wahrscheinlichkeit für eine Re-Identifikation signifikant, denn so ist es möglich, die Vermutung, dass eine potentiell identifizierte Person tatsächlich vor Ort gewählt hat, zu bekräftigen.

Jedenfalls wird hier klar, dass die herrschenden Anonymisierungskonzepte mindestens veraltet sind. Bereits jetzt werden Programme zur Verarbeitung von natürlicher Sprache dazu genutzt erfolgreich Anonymisierungsprozesse umzukehren (Deuber/Keuchen 2023). Das

dürfte ebenfalls dazu führen, dass sich der Pool an möglichen (technisch ungeschulten) Angreifern in Zukunft stark vergrößert.

Was den heutigen Stand im Kontext der Weitergabe der 73 Millionen anonymisierten Datensätze zur Krankengeschichte der gesetzlich versicherten für die Forschung angeht (Beuth 2022), bleibt zu sagen, dass der Prozess gegenwärtig pausiert ist (Kurz 2023). Zwar hat das Forschungsdatenzentrum des Bundes die Daten vom Bund der Krankenkassen bereits erhalten, allerdings kann die Ausgabe der Daten an die Forschung noch nicht vollzogen werden.

Das liegt mitnichten an einem aufgetauchten Störgefühl im Kontext des persönlichen Datenschutzes. Vielmehr wurde im Eilverfahren von Constanze Kurz vor dem Sozialgericht Berlin am 15.02.23 bekannt, dass das Forschungsdatenzentrum kein IT-Sicherheitskonzept habe und dass man gegenwärtig auf der Suche nach einem neuen IT-Dienstleister sei (ebd.; Gesellschaft für Freiheitsrechte 2023).

5 Quellenverzeichnis

Appenzeller, Arno/ Orak, Berna (2024): Das Re-Identifikationsrisiko bei der Weitergabe von Gesundheitsdaten. In: Datenschutz und Datensicherheit (Heft 5, 2024). S. 277-283.

Art.-29-Datenschutzgruppe (2014): Stellungnahme 5/2014 zu Anonymisierungstechniken. Bei Onlinequelle: https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_de.pdf (Abrufdatum: 16.9.24).

Beuth, Patrick (2022): Bürgerrechtler klagen gegen Weitergabe von Gesundheitsdaten. In: Der Spiegel. Bei Onlinequelle: https://www.spiegel.de/netzwelt/netzpolitik/buergerrechtler-klagen-gegen-weitergabe-von-gesundheitsdaten-a-9e2d37e3-857a-4209-9015-98f1f05d0bcd# (Abrufdatum: 10.9.24).

Bundesgesetzblatt (1977): Teil I, Ausgegeben zu Bonn am 1. Februar 1977, Nr.7. S. 201-215. Onlinequelle bei: https://www.bgbl.de/xaver/bgbl/start.xav?start=%2F%2F*%5B%40attr_id%3D%27bgbl177007.pdf%27%5D#__bgbl__%2F%2F*%5B%40attr_id%3D%27bgbl177007.pdf%27%5D__1722688863007 (Abrufdatum: 10.9.24).

Bundesgesetzblatt (2003): Teil I, Ausgegeben zu Bonn am 24. Januar 2003, Nr.3. S. 66- 88. Onlinequelle bei: https://www.bgbl.de/xaver/bgbl/start.xav#__bgbl__%2F%2F*%5B%40attr_id%3D%27bgbl103s0066.pdf%27%5D__1722847268519 (Abrufdatum: 11.9.24).

Bundesverfassungsgericht (1983): BVerfG: Verfassungsrechtliche Überprüfung des Volkszählungsgesetzes 1983. In: Neue Juristische Wochenschrift. S. 419-428. Beck-online.

Bundeswahlleiterein (2024): Information für Wählende. Bei Onlinequelle: https://www.bundeswahlleiterin.de/europawahlen/2024/informationen-waehler.html (Abrufdatum: 17.9.24).

Coucil of Europe (1981): Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data. 28.01.1981. Onlinequelle bei: https://rm.coe.int/1680078b37 (Abrufdatum: 10.9.24).

Deuber, Dominic/ Keuchen, Michael (2023): Gerichtsentscheidungen de-anonymisiert. In: Rechts EMPIRE. Bei Onlinequelle: https://rechtsempirie.de/post/gerichtsentscheidungen-de-anonymisiert/ (Abrufdatum 15.9.24).

Deutscher Ethikrat (2017): Big Data und Gesundheit – Datensouveränität als informationelle Freiheitsgestaltung. Onlinequelle bei: https://www.ethikrat.org/fileadmin/Publikationen/Stellungnahmen/deutsch/stellungnahme-big-data-und-gesundheit.pdf (Abrufdatum: 12.9.24).

Esayas, Samson (2015): The Role of Anonymisation and Pseudonymisation Under the EU Data Privacy Rules: Beyond the 'All or Nothing' Approach. In: European Journal of Law and Technology, Vol 6 (No. 2).

Europäischer Gerichtshof (2017): EuG: Bestimmung des Personenbezugs von Daten. In: Zeitschrift für Datenschutz. S. 24-29. Beck-online.

Geobasis NRW (2024): Digitale Orthophotos. WMS DOP in Farbe: https://www.wms.nrw.de/geobasis/wms_nw_dop. Bei Onlinequelle: https://www.bezreg-koeln.nrw.de/geobasis-nrw/produkte-und-dienste/luftbild-und-satellitenbildinformationen/aktuelle-luftbild-und-0 (Abrufdatum 18.9.24).

Gericht der europäischen Union (2023): EuGH: Speicherung von IP-Adressen beim Besuch einer Internetseite. In: Zeitschrift für Datenschutz. S. 399-404. Beck-online.

Gesellschaft für Freiheitsrechte (2023): Datenleak verhindern: Gesundheits-Datenbank von 73 Millionen gesetzlich Versicherter. Bei Onlinequelle: https://freiheitsrechte.org/themen/freiheit-im-digitalen/gesundheitsdaten (Abrufdatum: 18.9.24).

Goltz, Johannes/ Grunert, Hannes/ Heuer, Andreas (2017): De-Anonymisierungsverfahren: Kategorisierung und Anwendung für Datenbankanfragen. In: Leyer, M. (Hrsg.): Proceedings of the LWDA 2017 Workshops: KDML, FGWM, IR, and FGDB. Rostock, Germany, 11.-13. September 2017. Bei Onlinequelle: https://ceur-ws.org/Vol-1917/paper22.pdf (Abrufdatum: 14.9.24).

Hafner, Katie (2006): And if You Liked the movie, a Netflix Contest May Reward You Handsomely. In: The New York Times. Bei Onlinequelle: https://www.nytimes.com/2006/10/02/technology/and-if-you-liked-the-movie-a-netflix-contest-may-reward-you.html (Abrufdatum 20.9.24).

Hamacher, Kay/ Kussel, Tobias/ von Landesberger, Tatjana/ Baumgartl, Tom/ Höhn, Markus/ Scheithauer, Simone/ Marschollek, Michael/ Wulff, Antje (2022): Fallzahlen, Re-Identifikation und der technische Datenschutz. In: Datenschutz und Datensicherheit (Heft 3, 2022). S. 143-148.

Hölzel, Julian (2018): Anonymisierungstechniken und das Datenschutzrecht. In: Datenschutz und Datensicherheit (Volume 42, 2018). S. 502-509. https://doi.org/10.1007/s11623-018-0988-z.

Huschka, Denis/ Oellers, Claudia (2013): Einführung: Warum qualtitative Daten und ihre Sekundäranalyse wichtig sind. In: Huschka, Denis/ Knoblauch, Hubert/ Oellers, Claudia/ Solga, Heike (Hrsg.): Forschungsinfrastrukturen für die qualitative Sozialforschung. S. 9-18. Berlin: Scivero Verlag.

Kinder-Kurlanda, Katharina/ Watteler, Oliver (2015): Hinweise zum Datenschutz: Rechtlicher Rahmen und Maßnahmen zur datenschutzgerechten Archivierung sozialwissenschaftlicher Forschungsdaten. In: GESIS Papers, 2015/01. https://doi.org/10.21241/ssoar.43183.

Kneuper, Ralf (2022): Anonymisierte Daten brauchen keinen Datenschutz – wirklich nicht? In: Friedewald, Michael/ Kreutzer, Michael/ Hanse, Marit (Hrsg.): Selbstbestimmung, Privatheit und Datenschutz. S. 171-190. Wiesbaden: Springer Vieweg. https://doi.org/10.1007/978-3-658-33306-5_9.

Krebs, Hein-Adalbert/ Hagenweiler, Patricia (2022): Datenanonymisierung im Kontext von Künstlicher Intelligenz und Big Data. Wiesbaden: Springer Vieweg. https://doi.org/10.1007/978-3-658-37588-1.

Kretzer, Susanne (2013): Arbeitspapier zur Konzeptentwicklung der Anonymisierungs-/Pseudonymisierung in Qualiservice. In: Social Science Open Access Repository. https://nbn-resolving.org/urn:nbn:de:0168-ssoar-47605-2.

Kurz, Constanze (2023): Constanze Kurz über Chatkontrolle, Palantir & Künstliche Intelligenz. In: Jung, Thilo (Hrsg.) Jung & Naiv: Folge 680. Bei Onlinequelle: https://www.youtube.com/watch?v=Cb3mLfidgPY (Abrufdatum: 19.9.24).

Narayanan, Arvind/ Shmatikov, Vitaly (2007): Robust De-anonymization of Large Datasets (How to Break Anonymity of the Netflix Prize Dataset. In: Computer Science – Cryptography and Security (22. Nov 2007). https://doi.org/10.48550/arXiv.cs/0610105.

Pohle, Jörg/ Hölzel, Julian (2020): Anonymisierung aus Sicht des Datenschutzes und des Datenschutzrechts. In: Stellungnahme 29.06.2020 zum Konsultationsverfahren des BfDI zur Anonymisierung unter der DSGVO unter besonderer Berücksichtigung der TK-Branche, Alexander von Humboldt Institut für Internet und Gesellschaft. Onlinequelle bei: https://www.bfdi.bund.de/SharedDocs/Downloads/DE/Konsultationsverfahren/1_Anonymisierung/Stellungnahmen/Alexander-von-Humboldt-Institut.pdf?__blob=publicationFile&v=5 (Abrufdatum: 10.9.24).

Robinson, Dan (2022): Web daddy Tim Berners-Lee on privacy, data sharing, and the web's future. In: The Register. Bei Onlinequelle: https://www.theregister.com/2022/01/20/tim_bernerslee/ (Abrufdatum: 20.9.24).

Rocher, Luc/ Hendrick, Julien M./ de Montjoye, Yves-Alexandre (2019): Estimating the success of re-identifications in incomplete datasets using generative models. In: Nature Commun (Volume 10). https://doi.org/10.1038/s41467-019-10933-3

Schild, Hans Hermann (2024): DS-GVO Art.4 Rn. 14-21d. In: Wolff, Heinrich Amadeus/ Brink, Stefan/ Ungern-Sternberg, Antje (Hrsg.): BeckOK Datenschutzrecht. 48. Edition (Stand: 01.05.2024). München: C.H. Beck.

Schwartmann, Rolf/ Jaspers, Andreas/ Lepperhoff, Niels/ Weiß, Steffen/ Meier, Michael (2022): Praxisleitfaden zum Anonymisieren personenbezogener Daten. Bei Onlinequelle: https://stiftungdatenschutz.org/fileadmin/Redaktion/Dokumente/Anonymisierung_personenbezogener_Daten/SDS_Studie_Praxisleitfaden-Anonymisieren-Web_01.pdf (Abrufdatum: 13.9.24).

Snowden, Edward (2019): Edward Snowden on GDPR and data security. In: Verdict 3ncryptm, Winter 2019. Issue 11. Bei Onlinequelle: https://verdict-encrypt.nridigital.com/verdict_encrypt_winter19/edward_snowden_data_protection_collection_gdpr (Abrufdatum 11.9.24).

Sorge, Christoph (2013): Empirische Forschung im technischen Datenschutz: ein juristisches Problem? In: Abstraktion und Applikation: Tagungsband des 16. Internationalen Rechtsinformatik Symposions / Abstraction and Application: Proceedings of the 16th International Legal Informatics Symposium. https://publications.cispa.saarland/id/eprint/377.

Stadt Bottrop (2024a): Stadt Bottrop Europawahl 09.06.2024 – Ratswahlbezirke. Bei Onlinequelle: https://wahl.krzn.de/ew2024/wep040/navi/040-299-EW-KW.html (Abrufdatum: 17.9.24).

Stadt Bottrop (2024b): Stadt Bottrop Europawahl 09.06.2024 – Stadtbezirke. Bei Onlinequelle: https://wahl.krzn.de/ew2024/wep040/navi/040-299-EW-BEZ.html (Abrufdatum: 17.9.24).

Stadt Bottrop (2024c): Stadt Bottrop Europawahl 09.06.2024 – 1 Projektraum Hansastraße. Bei Onlinequelle: https://wahl.krzn.de/ew2024/wep040/erg/040-299-EW-s1.html (Abrufdatum: 17.9.24).

Stadt Bottrop (2024d): Bevölkerungszahlen nach Statistischen Bezirken (Stand 31.12.2023). Bei Onlinequelle: https://www.bottrop.de/daten-karten/statistik/einwohnerentwicklung.php (Abrufdatum: 17.9.24).

Stadt Bottrop (2024e): Stadtprofil (2023). Bei Onlinequelle: https://www.bottrop.de/daten-karten/statistik/stadtprofil.php (Abrufdatum: 17.9.24).

Stadt Bottrop (2024f): Statistische Bezirke in Bottrop. Bei Onlinequelle: https://www.offenesdatenportal.de/dataset/statistische-bezirke (Abrufdatum: 17.9.24).

Stadt Bottrop (2024g): Kommunalwahlbezirke der Stadt Bottrop. Bei Onlinequelle: https://www.offenesdatenportal.de/dataset/kommunalwahlbezirke-der-stadt-bottrop (Abrufdatum: 17.9.24).

Stadt Bottrop (2024h): Stadt Bottrop Europawahl 09.06.2024 – 6 Fuhlenbrock Heide. Bei Onlinequelle: https://wahl.krzn.de/ew2024/wep040/erg/040-299-EW-KW-d6.html (Abrufdatum: 17.9.24).

Stadt Bottrop (2024i): B) Wahlvorschläge für die Wahl in den Kommunalwahlbezirken. Bei Onlinequelle: https://www.bottrop.de/downloads/downloads/rathaus/amtliche/Wahlvorschlaege-KWB.pdf (Abrufdatum: 17.9.24).

Sweeney, Latanya (2002): k-anonymity: a model for protecting privacy. In: International Journal on Uncertainty, Fuzziness and Knowledge-based Systems. Volume 10 (Issue 5). S. 557-570.

Ulbricht, Carsten (2015): Anonymisierung und Pseudonymisierung; Verschlüsselung. In Dorschel, Joachim (Hrsg.): Praxishandbuch Big Data. Wiesbaden: Springer Gabler. https://doi.org/10.1007/978-3-658-07289-6.

Universität Washington (2019): Data Anonymization and De-Identification Challenges and Options - August 2019. Bei Onlinequelle: https://privacy.uw.edu/wp-content/uploads/sites/7/2021/03/DataAnonymization_Aug2019.pdf (Abrufdatum: 16.9.24).

Vokinger, Kerstin Noëlle/ Muehlematter, Urs J. (2019): Re-Identifikation von Gerichtsurteilen durch «Linkage» von Daten(banken). Eine empirische Analyse anhand von Bundesgerichtsbeschwerden gegen (Preisfestsetzungs-)Verfügungen von Arzneimitteln. In: Jusletter, (02.09.2019). https://doi.org/10.5167/uzh-202863.

Warren, Samuel D./ Brandeis, Louis D. (1890): The Right to Privacy. In: Harvard Law Review, Vol. 4, No. 5, S. 193-220. https://www.jstor.org/stable/1321160.

Watteler, Oliver/ Kinder-Kurlanda, Katharina (2015): Anonymisierung und sicherer Umgang mit Forschungsdaten in der empirischen Sozialforschung. In: Datenschutz und Datensicherheit (Heft 8). S. 515-519.

Wilhelm, Sebastian/ Folz, Jakob/ Wahl, Florian (2023): Open Personal Data: Anonymisierung im Spannungsfeld zwischen Informationsgehalt und Robustheit. In: Friedewald, Michael/ Karaboga, Murat (Hrsg.): Data Sharing: Datenkapitalismus by Default? Posterproceedings – Forum Privatheit 2023. Forum Privatheit und selbstbestimmtes Leben in der digitalen Welt, Karlsruhe: Fraunhofer ISI.

Eidesstattliche Erklärung &
Einwilligungserklärung Nutzung von Plagiatssoftware

Name:	Bleske	Studiengang:	SoSe 2024 MPA
Vorname:	Benjamin	Mtk.-Nr.:	36104066
Geb.-Ort:	Witten	Geb.-Datum:	11.02.1997

Mir ist bekannt, dass bei meiner Arbeit eine Prüfung auf nicht kenntlich gemachte übernommene Textpassagen und sonstige Quellen stattfinden kann (vgl. u.a. § 16 Abs. 7 der Allgemeinen Bestimmungen für Fachprüfungsordnungen mit den Abschlüssen Bachelor und Master der Universität Kassel). Ich stimme zu, dass dafür gegebenenfalls ein Upload auf eine externe Datenbank des jeweiligen Software-Anbieters erfolgt und die Arbeit dafür auch gespeichert wird, sofern meine Arbeit dafür vorab ausreichend anonymisiert wird (i.d.R. genügt dafür die Entfernung des Deckblatts und der Unterschriftenseite). Ich stimme ebenfalls zu, dass zukünftig umgekehrt auch andere Arbeiten auf Plagiate aus meiner anonymisierten Arbeit überprüft werden.

Ich versichere hiermit, dass ich meine Hausarbeit, Das Risiko der Re-Identifikation von anonymisierten Daten mit kritischer Würdigung selbständig und ohne fremde Hilfe angefertigt habe. Alle von anderen Autoren wörtlich oder sinngemäß übernommenen Stellen sind entsprechend gekennzeichnet.

Mir ist bewusst, dass bei einem Verstoß gegen obige Erklärung nicht nur die betreffende Prüfungsleistung mit der Note – 5,0 – gewertet wird, sondern auch eine Exmatrikulation erfolgen kann.

Der Prüfungsausschuss entscheidet im Einzelfall.

Bottrop, 27.09.2024	Benjamin Bleske, Bleske
Ort, Datum	Unterschrift