Anonymisierung von Gerichtsurteilen – Eine wesentliche Voraussetzung für E-Justice

unedited manuscript — please refer to the published paper for reliable citations

Prof. Dr. Axel Adrian, Prof. Dr. Stefan Evert, Michael Keuchen, Philipp Heinrich, Natalie Dykes

In Deutschland gibt es bereits viele Einzelanwendungen im Bereich Legal-Tech. Es fehlen aber Predictive Tools wie in den USA oder automatisch erstellte Urteilsentwürfe, weil nicht genügend Urteile als Trainingsdaten für subsymbolische KI-Verfahren zur Verfügung stehen. Bisher werden wohl nicht einmal 2 % aller Urteile veröffentlicht, da hierfür eine Anonymisierung erforderlich ist, die immer noch vollständig manuell erfolgt. Wir arbeiten im Rahmen eines Forschungsprojektes mit dem Bayerischen Staatsministerium der Justiz an rechtlichen und technischen Fragen zur Möglichkeit einer automatischen Anonymisierung von Urteilen, insbesondere mit Hilfe korpuslinguistischer Verfahren.

1. Einleitung

Im vorliegenden Beitrag werden einige Gedanken, sowohl aus der Rechtswissenschaft als auch aus der Korpus- und Computerlinguistik vorgestellt, um die Notwendigkeit einer möglichst automatischen Anonymisierung von Gerichtsurteilen, sowie deren rechtliche Grundlagen und technische Voraussetzungen zu erörtern.

1.1. Beispiele fortschreitender Digitalisierung im Bereich E-Government und E-Justice

Im Zusammenhang mit E-Government und E-Justice gibt es in Deutschland bereits für spezielle Anwendungen zahlreiche Legal-Tech-Tools.[1] So werden nach internen Angaben allein durch die Bayerische Finanzverwaltung von ca. 80 Millionen Steuerbescheiden schon fast 10 % ganz ohne menschliches Zutun erlassen, indem Expertensysteme zum Einsatz kommen. Städte setzen bereits Onlineplattformen zur Serviceverbesserung ein. Es werden Möglichkeiten der Erforschung einer automatischen Erstellung öffentlich-rechtlicher Verwaltungsakte diskutiert. Sofern es um öffentliches Baurecht geht, müssten dazu maschinenlesbare Bauanträge und Baugenehmigungen verarbeitet werden können, d.h. es geht nicht nur um natürliche Sprache, sondern auch um Pläne, also Bilder, die durch Maschinen zu verarbeiten wären.

Justiz und Notariat führten im Januar 2007 flächendeckend in Deutschland den elektronischen Rechtsverkehr zwischen Notariat und Handelsregister, und in der Folge auch mit verschiedenen anderen Kommunikationspartnern wie Grundbuchamt und diversen Behörden ein. Seit Jahren werden also z.B. alle Handelsregisteranmeldungen, manche Grundbuchanträge und Vorkaufsrechtsanfragen nurmehr elektronisch erledigt. Hierzu wird u.a. XNotar verwendet, eine digitale „Toolbox“ für verschiedene notarielle Aufgaben, wie das besondere elektronische Notarpostfach, als Nachfolger des bisherigen Elektronischen Gerichts- und Verwaltungspostfaches, die Verzeichnisse der Bundesnotarkammer (Notarverzeichnis, Notarportal, Stammdatenverzeichnis, Standesamtsverzeichnis und das Gerichtsverzeichnis) sowie der XKR-Kostenrechner. Wie auch sonst bei Legal-Tech üblich, werden rechtliche Prozesse immer schrittweise dort, wo es möglich ist, digitalisiert. So erreichte der elektronische Rechtsverkehrs in der notariellen Praxis eine beachtliche Leistungsfähigkeit.[2]

---

[1] Der Begriff „Legal-Tech“ ist in Deutschland ungefähr seit 2015 in aller Munde, siehe *HARTUNG*, Gedanken zu Legal Tech und Digitalisierung. In: Hartung/Bues/Halbleib (Hrsg.), Legal Tech, Beck, München 2018, S. 5 ff. Man spricht insoweit auch gerne von einem „Hype“.

[2] Dazu *BÜTTNER/FROHN/SEEBACH*, Elektronischer Rechtsverkehr und Informationstechnologie im Notariat, Beck, München 2019, S. 371, 355 f., 399, 431: Das Zentrale Vorsorgeregister (ZVR) dient z.B. dem besseren Auffinden von Vorsorgevollmachten und Betreuungsverfügungen. Durch die Justiz wird das ZVR ca. 20.000 Mal im Monat abgefragt. Bei rund 9,5 % der Fälle kann mindestens eine passende Eintragung beauskunftet werden. Dabei ist festzustellen, dass fast 90 % der Registrierungen im ZVR aus Notariaten stammen. Über das Zentrale

Weiterhin besteht eine große Herausforderung, insbesondere in unterschiedlichen Konzepten und Strategien der Bundesländer, die Schnittstellen und „Medienbrüche" verursachen. So sind die Unterschiede der in der Justiz genutzten Software – einerseits SolumStar (in 14 Bundesländern) und andererseits Folia/EGB (in Baden-Württemberg und Schleswig-Holstein) – für eine weitere Digitalisierung der Prozesse hinderlich.[3] Auch sollte man den vollständigen (semantischen) Inhalt der Registeranmeldungen, Grundbuchanträge, etc. und natürlich der Grundbücher selbst maschinell verarbeiten können. Die derzeit verwendeten PDF/A-Dokumente und XML-Strukturdaten sind aber für eine maschinelle Sprachverarbeitung (engl. *natural language processing*, NLP) nicht ausreichend; vielmehr werden maschinelesbare Repräsentationen natürlicher Sprache benötigt.

Justiz und Anwaltschaft haben im September 2018 nun flächendeckend jedenfalls mit „passiver Nutzungspflicht" das besondere elektronische Anwaltspostfach (beA) eingeführt. Auch hier bleibt kritisch zu fragen, inwieweit die Formate der eingereichten Schriftsätze tatsächlich in der Form maschinenlesbar sind, dass darauf aufbauend komplexere Legal-Tech-Tools (mit semantischer Textverarbeitung) entwickelt werden können.

1.2. Bedeutung anonymisierter Urteile als Trainingsdaten zur Fortentwicklung von E-Justice

Obwohl nun über das beA entsprechende Schriftsätze und auch aufgrund der in der Justiz eingesetzten Software die daraus resultierenden Urteile bereits digital vorliegen, können diese (noch) nicht in ausreichender Zahl für NLP-Analysen sowie als Trainingsdaten[4] für maschinelle Lernverfahren (engl. *machine learning*, ML) genutzt werden, was aber eine wesentliche Voraussetzung für die Entwicklung weiterführender Legal-Tech-Tools wäre. Daher fehlen in Deutschland bis heute z.B. gut funktionierende Predictive Tools, wie diese in den USA bereits erfolgreich eingesetzt werden[5], oder gar ein Werkzeug zur automatischen Erstellung von Urteilsentwürfen auf Grundlage bisheriger Gerichtsentscheidungen.[6]

Dass nicht genügend Gerichtsurteile als Trainingsdaten zur Verfügung stehen[7] überrascht zunächst, weil in Deutschland die Rechtsprechung seit Jahren eine Veröffentlichungspflicht von Entscheidungen aus dem Rechtsstaatsgebot einschließlich der Justizgewährungspflicht, dem Demokratiegebot und dem Grundsatz der Gewaltenteilung judiziert.[8] Das Problem liegt jedoch darin, dass die Urteile durch die Justizverwaltungen nur anonymisiert veröffentlicht werden dürfen. Da die Anonymisierung derzeit bundeslandspezifisch auf verschiedenen Rechtsgrundlagen umgesetzt wird, technisch nicht in einheitlicher Form und insbesondere noch vollständig manuell erfolgt, sind wohl nicht einmal 2 % aller Gerichtsurteile in Deutschland veröffentlicht.[9]

Testamentsregister (ZTR) soll sichergestellt werden, dass Verfügungen von Todes wegen aufgefunden und vollzogen werden. Bis zum 01. Januar 2012 „erfolgte die Benachrichtigung in Nachlasssachen mit papiergebundenen Verwahrungsnachrichten, die in rund 5.000 Testamentsverzeichnissen bei den (…) Standesämtern (…) dezentral geführt wurden." Die Bundesnotarkammer hat „innerhalb von rund drei Jahren" in einem der „größten Projekte zur Verwaltungsmodernisierung" ca. 18,4 Mio. solcher Verwahrungsnachrichten digitalisiert und in das ZTR digital überführt. Mit diesem System werden heute pro Werktag ca. 4.000 Sterbefälle bearbeitet und daraufhin untersucht, ob sich unter den mittlerweile 20 Mio. hinterlegten Registrierungen ein korrespondierender Datensatz befindet.

3 *ibid.*, S. 542 f.

4 Bereits im Beitrag *ADRIAN,* Der Richterautomat ist möglich – Semantik ist nur eine Illusion?, RECHTSTHEORIE 2017, S. 77 ff. wurde gefordert nicht nur Urteile als solche, sondern grundsätzlich auch alle Schriftsätze, die historisch zum Erlass des jeweiligen Urteils führten, in die Datenbasis von Legal-Tech-Tools miteinzubeziehen, um stabile Ergebnisse zu erhalten. Dies ist auch mit Überlegungen aus der Methodenlehre, der Wissenschaftstheorie und der Rechtsphilosophie zu begründen, was in *ders.,* Grundprobleme einer juristischen (gemeinschaftsrechtlichen) Methodenlehre, Berlin 2009, *ders.,* Grundzüge einer allgemeinen Wissenschaftstheorie auch für Juristen, Berlin 2014, *ders.,* Wie wissenschaftlich ist die Rechtswissenschaft?, RECHTSTHEORIE 2010, S. 521 ff., ausführlich dargelegt wurde.

5 *VOGL,* Changes in the US Legal Market Driven by Big Data/Predictive Analytics and Legal Platforms. In: Hartung/Bues/Halbleib (Hrsg.), Legal Tech, Beck, München 2018, S. 53 ff; *ASHLEY,* Artificial Intelligence and legal Analytics, Cambridge University Press, 2017, S. 107, 109, 111, 125 f.; *FREUDENTHALER*, Case-based-Reasoning (CBR): Grundlagen und ausgewählte Anwendungsbeispiele des fallbasierten Schließens, AkademikerVerlag, Saarbrücken 2008; *BENCH-CAPON/SARTOR*, A Model of Legal Reasoning with Cases Incorporating Theories and Values, Artificial Intelligence 150, 2003, S. 97 ff.

6 Siehe zur prinzipiellen Möglichkeit dieser Vision *ADRIAN,* Der Richterautomat ist möglich – Semantik ist nur eine Illusion?, RECHTSTHEORIE 2017, S. 77 ff., wobei die juristische Methodenlehre nicht unmittelbar als Vorbild dienen kann, was in *ADRIAN,* Juristische Methodenlehre – Ein Vorbild für verantwortungsvolle Digitalisierung?. In: Schweighofer/Hötzendorfer/Kummer/Saarenpää (Hrsg.), Verantwortungsbewusste Digitalisierung, Tagungsband des 23. Internationalen Rechtsinformatik Symposiums IRIS 2020, Bern 2020, S. 41-48 dargelegt wurde.

7 *FRIES*, Rechtsdienstleistung durch KI. In: Kaulartz/Braegelmann (Hrsg.), Rechtshandbuch Artificial Intelligence und Machine Learning, Vahlen, München 2020, S. 655 f.

8 BVerwG 26. Februar 1997 - 6 C 3.96; BGH 05. April 2017 - IV AR (VZ) 2/16.

9 „Weniger als 2 %" werden typischerweise genannt, wenn *COUPETTE/FLECKNER*, Quantitative Rechtswissenschaft, JZ 2018, S. 379 (S. 381) auch nur von einer sehr kleinen Teilmenge sprechen, oder *HARTUNG* mitteilt, dass 99 % aller Urteile nicht in digitaler Form verfügbar sind:

Insbesondere mit Hilfe korpus- und computerlinguistischer Verfahren arbeiten wir im Rahmen eines Forschungsprojektes mit dem Bayerischen Staatsministerium der Justiz an rechtlichen und technischen Fragen zur Möglichkeit einer im Idealfall vollautomatischen[10] Anonymisierung von Urteilen. Solange anonymisierte Urteile nicht in ausreichender Zahl als Trainingsdaten für subsymbolische KI-Verfahren verfügbar sind, wird der Rechtsstandort Deutschland im Vergleich zu den USA ins Hintertreffen geraten.

2. Rechtliche Herausforderungen bei der Anonymisierung von juristischen Texten

Bei der Anonymisierung von juristischen Texten stellen sich einige oft unterschätze Herausforderungen. Zunächst ist es im Vergleich zu strukturierten Mikrodaten, die meist in tabellarischer Form vorliegen, nicht trivial, die zu anonymisierenden sensitiven Informationen in einem Fließtext zu lokalisieren.[11] Dafür muss der Mensch oder die Maschine erkennen, in welchen Wörtern kritische Informationen enthalten sind. Dies gestaltet sich besonders schwierig, da an praktisch allen Stellen des Textes solche Informationen auftauchen können oder diese erst im Zusammenspiel mehrerer Textstellen entstehen. Um diese Aufgabe zu bewältigen, muss man sich fragen, welche Informationen als kritisch einzustufen sind, um diese einer Risikobewertung zu unterziehen und schließlich entscheiden zu können, in welcher Weise sie behandelt werden müssen. Insgesamt zeigt sich bei der Anonymisierung von Fließtext eine besondere Schwierigkeit, da nicht nur ein angemessener Schutz vor De-Anonymisierung bestehen muss, sondern auch die Syntax und Semantik des Textes sowie die Relationen zwischen Personen und andere juristisch relevante Informationen hinreichend erhalten bleiben müssen.

2.1. Anonymisierung – Vom Volkszählungsgesetz zur modernen Kryptografie

Die Bedeutung des altgriechischen Worts *anónymos* ist „ohne Namen“. Doch die Persönlichkeit und besonders die Personenbeziehbarkeit ergeben sich nicht nur aus direkten Identifikatoren wie Name oder Adresse, sondern auch aus indirekten sowie kombinierten Identifikatoren, die insgesamt ein Individuum prägen. Diese Identifikatoren können aus zahlreichen unterschiedlichen statischen wie wandelbaren Merkmalen bestehen, wie Geschlecht, Beruf oder Gesundheitszustand, aber auch wesentlich fernliegenderen Informationen.[12] So kann bspw. ein Wohnort nicht nur über die Adresse dargestellt werden, sondern über deskriptive Merkmale, wie das (einzige) rote Haus in einem bestimmten kleinen Dorf. Hier lauert eine häufig unterschätzte Gefahr in Urteilen aufgrund von kostengünstigen Verknüpfungsmöglichkeiten[13] eine De-Anonymisierung herbeizuführen (vgl. zur Häufigkeit von indirekten Identifikatoren die ersten Projektergebnisse unter 3.3.).

Aus den komplexen Merkmalen formt sich das Recht auf informationelle Selbstbestimmung, welches das Bundesverfassungsgericht im Volkszählungsurteil aus dem allgemeinen Persönlichkeitsrecht in Art. 2 Abs. 2 GG i.V.m. der Menschenwürde in Art. 1 Abs. 1 GG entwickelt hat. Jedem Individuum steht es frei, grundsätzlich selbst über die Preisgabe und Verwendung seiner persönlichen Daten zu bestimmen und vor unbegrenzter Erhebung, Speicherung, Verwendung und Weitergabe geschützt zu sein.[14] Hintergrund der Verfassungsbeschwerden war das Volkszählungsgesetz, das dem Staat statistische Ergebnisse über gesellschaftliche und wirtschaftliche Entwicklungen zur Vorbereitung von Entscheidungen geben sollte. Doch durch die damals zunehmende technische Erfassung und Verknüpfbarkeit von Datenbeständen wuchsen die Gefahren von Re-Identifikation und Persönlichkeitsprofilbildung durch Kombination von Zusatzwissen und Datenbeständen. Kurzum, das Statistikgeheimnis aus § 11 des Bundesstatistikgesetz von 1980 und die judizierte Anonymität aus dem Mikrozensus-Beschluss[15] des Bundesverfassungsgerichts wurden als beeinträchtigt gesehen.

---

www.netzpiloten.de/werkzeuge-daten-gerichtsurteile (abgerufen am 26.10.2020). Siehe zu älteren Statistiken *KUNTZ*, Quantität gerichtlicher Entscheidungen als Qualitätskriterium juristischer Datenbanken, JurPC Web-Dok. 12/2006, Abs. 34: dort ist die Rede von 0,27 – 4,95 %.

[10] Bisherige Software Lösungen zur (semi)automatisierten Anonymisierung erfordern zumeist eine exakte manuelle Eingabe der kritischen Textinformation sowie des Pseudonyms. Hierzu mit Lösungsbeispiel: *DÉVAUD/KUMMER*, (Semi-)Automatische Anonymisierung von Entscheiden, Jusletter IT 23. Februar 2017, S. 2 ff.

[11] *WINTER/BATTIS/HALVANI*, Herausforderungen für die Anonymisierung von Daten, ZD 2019, S.489 (S. 490).

[12] Stellungnahme 4/2007 vom 20. Juni 2007 der nach Art. 29 der Richtlinie 95/46/EG eingesetzten Datenschutzgruppe, 01248/07/DE WP 136, S. 15.

[13] *SCHILD*, in: Beck`scher Online-Kommentar Datenschutzrecht, Wolff/Brink (Hrsg.), 34. Edition 01.11.2020, Grundlagen und bereichsspezifischer Datenschutz, Syst. E., Rn. 58.

[14] BVerfG 15. Dezember 1983, 1 BvR 209/83, Rn. 147.

[15] BVerfG, 16. Juli 1969, 1 BvL 19/63.

Dieser historische Ausflug zeigt, wo die Wiege des modernen Datenschutzrechts im Sinne einer DSGVO liegt. Viele Begriffe wie Anonymität, Personenbeziehbarkeit, Re-Identifizierbarkeit und deren Zusammenhänge bestehen nach der Grundkonzeption fort. Auf dieser Basis haben sich jedoch einerseits die Anforderungen an die Anonymität und die Anonymisierungstechniken weiterentwickelt, aber andererseits auch die Gefahren einer De-Anonymisierung nochmals verschärft.[16] Zwar lässt bspw. die moderne Kryptografie mittels Blockchain-Technologie unter bestimmten Rahmenbedingungen ein anonymes und zugleich transparentes Register zu.[17] Doch gleichzeitig existieren so viele Daten wie noch nie über uns alle in öffentlich zugänglichen staatlichen oder privaten Datenbanken. Diesbezüglich haben einige Re-Identifikationsexperimente gezeigt, wie anfällig die Anonymität ist und mit wie wenigen Merkmalen eine Person identifiziert werden kann.[18]

Diese Gegebenheiten und die Entwicklung immer schneller wachsender Datenbanken, mit einer immensen quantitativen wie qualitativen Dimension, sollten eine Parallele zum ursprünglichen Statistikrecht aufzeigen. Komplexeste Datenbanken der großen Internetkonzerne entsprechen mittlerweile einer interaktiven Gesellschaftsstatistik, mit der Möglichkeit über eine Vielzahl an Individuen Vorhersagen zu treffen. Wenn in Zukunft große Datenbanken aus Urteilen, Bescheiden, Verträgen, Urkunden usw. für Legal-Tech-Anwendungen entstehen, dann birgt dies dieselbe Gefahr. Wenn bereits die Justizstatistik mit Angaben über Verfahrensausgänge usw. dem strengen Statistikrecht unterliegt, so muss erst recht eine Datenbank mit den unzähligen Einzelentscheidungen – mit wesentlich mehr Merkmalen und einem deutlich höheren Risiko einer Re-Identifizierung – einem strengeren Reglungsregime unterliegen als bloß der DSGVO. Vor diesen Besonderheiten sind die tradierten Begriffe wie Anonymität, Personenbeziehbarkeit und Re-Identifizierung unter Berücksichtigung der aktuellen technischen Möglichkeiten einer interdisziplinären[19] Begriffsbestimmung zu unterziehen.

2.2. Rethinking Anonymität, Personenbeziehbarkeit und Re-Identifikation

Die Überlegungen zur Begriffsbestimmung der Anonymität reichen bereits Jahrzehnte zurück. Schnell wurde klar, dass es eine absolute Anonymität nach rein formalen (mathematischen) Kriterien nicht gibt.[20] Die zentrale Frage ist also, ab welchem (relativen) Grad von zu investierendem Re-Identifikationsaufwand ein Datensatz als anonym gelten kann (sog. faktische Anonymität). Dies zeigt, wie normativ der Begriff der Anonymität aufgeladen ist, und dass dieser erst mit Hilfe verschiedener normativer Kriterien definiert werden kann. Betrachtet man ErwGr. 26 DSGVO, so ist entscheidend, ob sich die Information auf eine identifizierte oder identifizierbare natürliche Person bezieht. Zur Feststellung der faktischen Anonymität, insbesondere für den schwierigeren Fall einer indirekten Identifizierbarkeit, sind alle objektiven Faktoren wie die Kosten der Re-Identifizierung und der dafür erforderliche Zeitaufwand zu berücksichtigen, wobei dafür die zum Zeitpunkt der Anonymisierung verfügbare Technologie und die technologischen Entwicklungen maßgeblich sind. Diese Bestimmung macht deutlich, dass nicht nur der verarbeitete Datensatz von Relevanz ist, sondern außerdem das bestehende Zusatzwissen, welches vom Angreifer zur De-Anonymisierung herangezogen wird. Dieses Zusatzwissen besteht aus eigenen, allgemein zugänglichen oder nach bestimmten Berechtigungen legal verfügbaren Datensätzen, welche

---

[16] *BOEHME-NEßLER*, Das Ende von der Anonymität, DuD 2016, S. 419 (S. 422); *HORNUNG/WAGNER*, Anonymisierung als datenschutzrelevante Verarbeitung?, ZD 2020, S. 223.

[17] Dazu und am Beispiel eines Blockchain-Grundbuchs: *KEUCHEN*, Grundbuch 4.0 – Folgt das Blockchain-Grundbuch dem Datenbankgrundbuch?, ZfIR 2020, S. 593 ff. Es zeigt sich, viele Anwendungen scheitern mangels Datenbestand.

[18] Der Re-Identifikationsversuch aus dem Jahr 2018 auf Basis eines statistischen Models von *Rocher/Hendrickx/Montjoye* zeigt, dass in jedem Datensatz, der 15 demographische Merkmale enthält, 99,98 % der Amerikaner/innen zutreffend re-identifiziert werden können. *ROCHER/HENDRICKX/DE MONTJOYE,* Estimating the success of re-identifications in incomplete datasets using generative models, Nature Communications 10, 2019, Nr 3069. In einem weiteren Re-Identifikationsexperiment von österreichischen Bundesgerichtsbeschwerden gegen (Preisfestsetzungs-)Verfügungen von Arzneimitteln durch „Linkage" von einer Urteilsdatenbank des Bundesverwaltungsgerichts und einer Datenbank des Bundesamtes für Gesundheit führte bei einer Stichprobe von 25 Entscheidungen zu einer Re-Identifikation von 84 % der dort inkludierten Arzneimittel und Zulassungsinhaberinnen mit verhältnismäßigen Aufwand. *VOKINGER/MÜHLEMATTER*, Re-Identifikation von Gerichtsurteilen durch «Linkage» von Daten(banken), Jusletter 2. September 2019, S. 16.

[19] Zu diesem Gedanken bereits *HAMMERBACHER*, Die zu fordernde „ausreichende" Anonymisierung von Datensätzen, DuD 1984, S. 181 (S. 182). Er verlangt eine Gesamtentscheidung aus mathematisch-theoretischer und normativ-juristischer Sicht bei der Bewertung der Anonymität.

[20] Nach ISO 29100:2011, 2.2 wird eine Anonymisierung beschrieben als: „process by which personally identifiable information (PII) is irreversibly altered in such a way that a PII principal can no longer be identified directly or indirectly, either by the PII controller alone or in collaboration with any other party".

der Angreifer nutzt, um Überschneidungen in den Merkmalen oder Merkmalsausprägungen zu entdecken, damit die Merkmalsträger durch den Ausschluss von Doppelgängern mit einer hinreichenden Sicherheit re-identifiziert werden.[21]

Die bisherige Darstellung zeigt, dass der Begriff der Anonymität von zahlreichen Wertungsentscheidungen abhängt und an vielen Stellen verschiedenste Wahrscheinlichkeitsentscheidungen getroffen werden müssen. Obwohl diese Einzelwahrscheinlichkeiten aktuell mangels statistischen Datenmaterials nicht bestimmt oder bestimmbar sind, erfolgt dennoch eine binäre Entscheidung darüber, ob das Datenschutzrecht Anwendung findet oder nicht. Für pseudo-anonyme Daten existiert kein rechtliches Zwischenstadium und damit können Schutzlücken entstehen, wenn aufgrund von einer geschickten Kombination (*cross-referencing*) von Daten ein Personenbezug herstellbar ist.[22] Damit sich diese Wahrscheinlichkeiten beziffern lassen, muss in Zukunft in interdisziplinären Projekten eine empirische Forschung stattfinden, um das Anonymity-Set der Merkmalsträger quantitativ wie qualitativ greifbar zu machen.[23] Dies ist notwendig, um eine realistische Einschätzung abzugeben, ab wann eine Zuordnung der Einzelangaben zu einer betroffenen Person nur mit einem unvernünftig großen Aufwand an Zeit, Kosten und Arbeitskraft möglich ist.[24] Das wird helfen, die relative Identifizierbarkeit sichtbar zu machen und festzulegen, welches Zusatzwissen der verarbeitenden Stelle noch zuzurechnen ist.[25]

2.3. Anonymisierung von Urteilen – Nicht nur eine Frage der DSGVO

2.3.1. Schutz juristischer Personen

In Urteilen sind im erheblichen Umfang sensible Informationen über juristische Personen enthalten. Diese lassen sich allein mit den Vorschriften des Datenschutzrechts nicht adäquat schützen, weshalb in diesem Kontext weitere Rechtspositionen neben dem Schutz des Rechts auf informationelle Selbstbestimmung zu beachten sind. So sind juristische Personen nach ErwGr. 14 S. 2 DSGVO ausgeschlossen. Zwar erfolgt in Deutschland kein nationaler Schutz über das BSDG, dennoch schützt die Rechtsordnung die juristischen Personen in ihrem Ruf in der Öffentlichkeit, wobei stellenweise sogar ein Unternehmenspersönlichkeitsrecht angenommen wird.[26] Weiterer Schutz erfolgt im Rahmen des Statistikrechts.[27] Ein Ländervergleich zur Schweiz zeigt, dass dort unmittelbar nach Art. 2 DSG juristische Personen in den Geltungsbereich fallen. In Österreich ist die Rechtslage seit der Umsetzung der DSGVO in das nationale Recht im Wandel. Bis zur Gesetzesnovelle am 24. Mai 2018 galt nach Art. 4 § 2 Nr. 1 und 3 DSG 2010 eine juristische Person als Betroffene, wird aber seitdem nicht mehr ausdrücklich vom Wortlaut umfasst. Dennoch bejaht die Datenschutzbehörde wegen des in Art. 1 § 1 Abs. 1 DSG und Art. 8 GRCh kodifizierten Grundrechts auf Datenschutz für „jedermann" einen Schutz für juristische Personen durch die einfachgesetzlichen Bestimmungen des DSG.[28]

Stets unter das Datenschutzrecht fallen in Deutschland die hinter der juristischen Person stehenden natürlichen Personen, insbesondere die Organwalter oder Angestellten.[29] Werden Informationen über juristische Personen preisgegeben, so können sich diese leicht auf die handelnden natürlichen Personen beziehen und auf diese durchschlagen.[30] Diese Problematik trennscharf zu erfassen – insbesondere mit

---

21 Definition nach *PAAß/WAUSCHKUHN*, Datenzugang, Datenschutz und Anonymisierung, R. Oldenbourg Verlag, München 1984, S. 3.

22 *KARG*, Anonymität, Pseudonyme und Personenbezug revisited?, DuD 2015, S. 520.

23 Ähnlich auch *HORNUNG/WAGNER*, Der schleichende Personenbezug, CR 2019, S. 565 (S. 574), welche auch weitere Konkretisierungen zu Identifizierbarkeit bei Big Data-Anwendungen und empirische Untersuchungen zur Problematik eines schleichenden Personenbezugs anregen. *COUPETTE/FLECKNER*, Quantitative Rechtswissenschaft, JZ 2018, S. 379 (S. 389), befürworten ebenso quantitative Studien in der Rechtswissenschaft.

24 Art. 29-Datenschutzgruppe, Stellungnahme 5/2014 zu Anonymisierungstechniken, 10. April 2014, Aktenzeichen 082914 0829/14/DE WP216, S. 6.

25 Für die herrschende relative Theorie: EuGH 19. Oktober 2016, C-582/14. Zur Berücksichtigung von Zusatz und Spezialwissen *SONNTAG*, Anonymisierung: Methoden und Zulässigkeit, Jusletter IT 25. Februar 2016, S. 6.

26 Zum guten Ruf aus dem Recht am eingerichteten und ausgeübten Gewerbebetrieb OLG München 27. Februar 2020, 29 U 2584/19. Ebenso dazu und zum Unternehmenspersönlichkeitsrecht BGH 14. Januar 2020, VI ZR 495/18.

27 So etwa ausdrücklich in Art. 2 Abs. 5 BayStatG. Ebenso *POPPENHÄNGER*, Die Übermittlung und Veröffentlichung statistischer Daten im Lichte des Rechts auf informationelle Selbstbestimmung, Berlin 1995, S. 41.

28 Datenschutzbehörde 25. Mai 2020, GZ. 2020-0.191.240, Rn. 49-65.

29 EuGH 09. November 2010, C-92/09 und C-93/09 ausdrücklich in Rn. 53 für den Fall, soweit der Name der juristischen Person eine oder mehrere natürliche Personen bestimmt.

30 *KARG*, in: Datenschutzrecht, Simitis/Hornung/Spiecker (Hrsg.), Nomos, Baden-Baden 2019, Art. 4 Nr. 1 DSGVO Rn. 44.

technischen Mitteln – stellt eine Herausforderung dar, weshalb darüber nachgedacht werden sollte, juristische Personen pauschal bei der Anonymisierung von Urteilen zu inkludieren.

Eine Heterogenität im Umgang mit der Anonymisierung juristischer Personen wegen einer angenommen Schutzwürdigkeit konnten wir auch bei einer Analyse der Rechtsprechungsdatenbank der Bayerischen Staatskanzlei „Bayern.Recht“[31] feststellen. Dort ergaben sich bei der Suche nach juristischen Personen und Unternehmen in den 6.310 erfassten Urteilen der ordentlichen Gerichtsbarkeit insgesamt 24.638 Vorkommen in 2.858 Urteilen. Die Behandlung war sehr unterschiedlich, von einer Anonymisierung mit Initialen in ca. 69 % der Fälle, mit „…“ in ca. 21 % der Fälle oder dem Belassen der Klarinformation in rund 10 % der Fälle. Dieses Bild zeigt, dass neben dem Datenschutzrecht ein Bedürfnis besteht, die Informationen über juristische Personen in Urteilen zu neutralisieren. Andererseits kommt diesen Beteiligten ein geringeres Schutzniveau zu, da sich diese bewusst in die Öffentlichkeit begeben, in besonderem Maße in der Sozialsphäre aktiv sind und wahrgenommen werden wollen. Deshalb sind meist Informationen über Marken nicht zu anonymisieren.[32]

2.3.2. Steuer-, Betriebs- und Geschäftsgeheimnis

Neben den einfachgesetzlichen Vorschriften des Datenschutzes bestehen weitere Normen, die einer vollständigen Veröffentlichung einer Entscheidung entgegenstehen. Exemplarisch wird auf das Steuergeheimnis nach § 30 AO eingegangen, welches die persönlichen und wirtschaftlichen Umstände eines anderen (juristische oder natürliche Person) sowie Betriebs- und Geschäftsgeheimnisse, die einem Amtsträger bekannt geworden sind, vor einer unbefugten Offenbarung oder Verwertung schützt. Deshalb unterblieb in der Steuerstrafsache gegen Uli Hoeneß im Urteil ein Teil der Feststellungen zum privaten und beruflichen Werdegang.[33] Andererseits finden sich in vielen finanzgerichtlichen Entscheidungen zahlreiche Einzelangaben zu Einkommen, Familienverhältnissen, Nebentätigkeiten, außergewöhnlichen Belastungen oder Grundstücksgrößen und lassen eine Identifizierbarkeit zu.[34] Auch wenn § 30 AO nur eine einfachgesetzliche Ausprägung von Art. 2 Abs. 1 i.V.m. Art. 1 Abs. 1 GG, 14 GG und ggf. Art. 19 Abs. 3 GG ist, so verdeutlicht dies die Diversität der Merkmale, die es zu schützen gilt und die in Kombination eine Einmaligkeit aufweisen und Rückschlüsse zulassen.[35]

2.4. Rechtszweigspezifische Veröffentlichungsregelungen und Anonymisierungskriterien

Festgehalten werden kann, dass mehr Unklarheiten bei der Anonymisierungswürdigkeit von Merkmalen aus rechtlicher wie empirischer Sicht bestehen, als dies gemeinhin angenommen wird. Daher erscheint es geboten, den Erlass von rechtszweigspezifischen Regelungen für die Veröffentlichung von Entscheidungen zu diskutieren, sowie Kriterienkataloge mit Risikogruppen für bestimmte Merkmale zu entwickeln.[36] Danach ließe sich ein abgestuftes System von Anonymisierungsprozessen und -anforderungen nach dem jeweiligen Gefahrenpotenzial festlegen, sodass im Regelfall in einer ersten Stufe nur eine Anonymisierung von typischen Einzelinformationen erfolgt, in der zweiten Stufe spezifische Informationen bis hin zu ganzen Textpassagen anonymisiert werden, wobei auf der letzten Stufe als Ultima Ratio eine Veröffentlichung gänzlich unterbleibt.[37]

Wenn zukünftig komplexere Legal-Tech-Anwendungen entstehen sollen, dann muss eine immense Datenbasis geschaffen werden, die solch differenzierten Regelungen zur Anonymisierung unterworfen ist.

---

[31] Abrufbar unter: www.gesetze-bayern.de. Die Analyse wurde im Mai 2020 durchgeführt.

[32] OLG Frankfurt 19. September 2019, 20 VA 21/17.

[33] LG München II 13. März 2014 - W5 KLs 68 Js 3284/13. Das trotz Anonymisierung eine Zusammenführung von Name und Aktenzeichen möglich ist, zeigt abermals die Schwierigkeit. Jedoch kann bei Verfahren mit Personen der Zeitgeschichte ein größeres berechtigtes Interesse der Öffentlichkeit bestehen. So auch *PUTZKE/ZENTHÖFER*, Der Anspruch auf Übermittlung von Abschriften strafgerichtlicher Entscheidungen, NJW 2015, S. 1777 (S. 1781).

[34] *HAUPT*, (Kein) Steuergeheimnis nach dem Finanzgerichtsprozess?, DStR 2014, S. 1025 (S. 1029).

[35] BVerfG 17. Juli 1984, 2 BvE 11/83 und 15/83.

[36] *HAUPT*, (Kein) Steuergeheimnis nach dem Finanzgerichtsprozess?, DStR 2014, S. 1025 (S. 1030); *KÖPFERL,* „Die Veröffentlichung unterbleibt im Hinblick auf das Steuergeheimnis“, ZIS 2015, S. 375 (S. 385).

[37] Einen Vorschlag für Anonymisierungsstufen macht *NÖHRE*, Anonymisierung und Neutralisierung von veröffentlichungswürdigen Gerichtsentscheidungen, MDR 2019, S. 136 (S. 138 ff.).

3. Forschungsprojekt zur automatischen Anonymisierung von Gerichtsurteilen

Um die aufgeworfenen rechtlichen und empirischen Fragestellungen anzugehen, forschen wir für das Bayerische Staatsministerium der Justiz an der automatischen Anonymisierung von Urteilen. Ein erstes Ziel des Projekts ist die Entwicklung und detaillierte Evaluation eines Software-Prototypen für die automatische Erkennung sensitiver Textstellen. Voraussetzung dafür ist eine manuelle Kennzeichnung dieser Stellen, die in Kategorien eingeteilt und bezüglich ihres Risikoniveaus bewertet werden. Im weiteren Projektverlauf soll auf dieser Basis die Möglichkeit einer zielgerichteten informationserhaltenden Anonymisierung untersucht werden.

3.1. Ausgangslage und Datenbasis des Projekts

Das Datenmaterial, das uns in der ersten Projektphase zur Verfügung steht, umfasst knapp 300 Urteile aus dem Rechtsgebiet Mietrecht. Dies entspricht einem Gesamtumfang von 650.000 Token[38] Fließtext; im Schnitt also ca. 2.300 Token pro Urteil. Die Länge variiert dabei stark: das kürzeste Urteil umfasst lediglich 427 Token, das längste über 11.000. Dies liegt insbesondere an der unterschiedlichen Länge des Tatbestands und der Entscheidungsgründe. Das strukturell über alle Urteile hinweg sehr ähnliche Rubrum umfasst typischerweise knapp 100 Token, die Rechtsbehelfsbelehrung ca. 330 Token.

Der fortlaufende und iterative Arbeitsablauf des bis 2022 dauernden Forschungsprojektes ist in Abb. 1 dargestellt. So erhoffen wir uns Erkenntnisse zur rechtlichen Dogmatik, der technischen Möglichkeiten und zu Fragen, wie Recht und Technik für eine möglichst automatische Anonymisierung zusammenwirken können.[39]

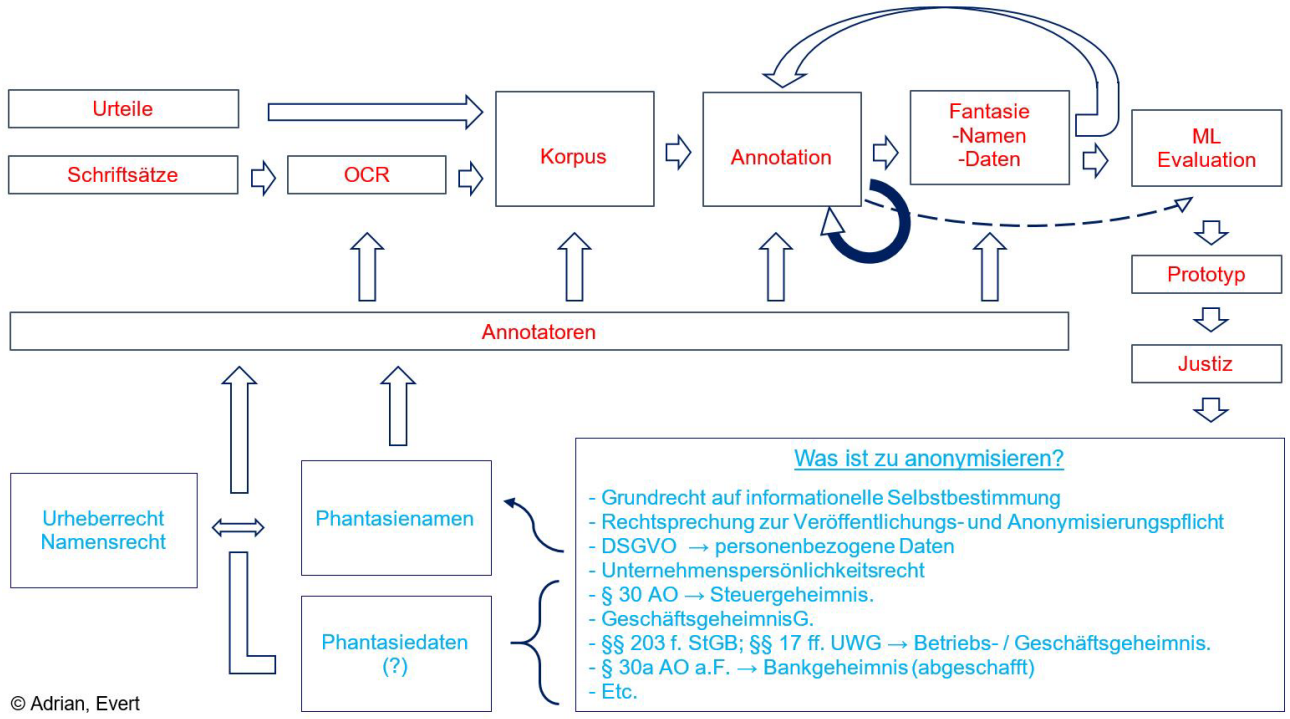

Abb. 1: Schematische Darstellung des Arbeitsablaufs des Forschungsprojekts

3.2. Vorgehensweise bei der Annotation sensitiver Textstellen

Für das Training eines maschinellen Lernverfahrens zur Erkennung sensitiver Textstellen sowie zu dessen Evaluation wird ein Goldstandard benötigt, d.h. ein Korpus von Urteilen, in dem die sensitiven Stellen manuell korrekt markiert wurden. Hierfür beschäftigen wir zahlreiche studentische Hilfskräfte, deren Aufgabe es ist, unabhängig voneinander diese Stellen zu identifizieren und zusätzlich den Grund für

---

[38] In der computerlinguistischen Verarbeitung werden Texte zunächst in Wörter und Satzzeichen als Grundeinheiten zerlegt, für die üblicherweise der Oberbegriff „Token" verwendet wird.

[39] Siehe hierzu umfassend die Vorlesung von *ADRIAN*, „Künstliche Intelligenz und juristisches Entscheiden", Einheit 6 (youtu.be/p-6JOIiBRbc) und 7 (youtu.be/yYcn2emJMos) aus dem Sommersemester 2020 an der FAU.

die Sensitivität der Stelle zu vermerken. Dieses Vorgehen wird in der Computerlinguistik als „Tagging" bezeichnet; die zuweisbaren Begründungskategorien werden in diesem Zusammenhang „Tags" genannt.

Die Hilfskräfte wurden in mehreren Probeläufen auf frei verfügbaren und synthetischen Datensätzen geschult. Parallel dazu wurde in mehreren Iterationen ein geeignetes Tagset entwickelt. Das Tagset mit derzeit ca. 20 Tags in sechs Oberkategorien und entsprechende Annotationsrichtlinien werden von uns stetig weiterentwickelt. Insbesondere soll noch geklärt werden, ob und inwieweit die erarbeiteten Grundlagen für die Anwendung auf Urteile aus wenigstens einem anderen Rechtsgebiet angepasst oder abgeändert werden müssen.

Die Übereinstimmung der Annotatoren untereinander (*inter-annotator agreement*) ist in den meisten Kategorien sehr hoch; insbesondere für einfach zu identifizierende Tags wie Datumsangaben und Namen. Lediglich bei den sonstigen identifizierenden Merkmalen findet sich ein hohes Maß an Subjektivität. Dies ist nicht weiter verwunderlich, da bspw. die Gefährlichkeit von indirekten Identifikatoren und deren Kombination, von den Annotatoren eigenständig nach Risikograden bewertet wird. Hierbei zeigt sich die bereits angesprochene Notwendigkeit einer empirischen Betrachtung dieser Merkmale. Da die Merkmale aber von mehreren Annotatoren eingeschätzt werden, lässt sich ein Durchschnittswert ermitteln. Schließlich wird das Erfordernis einer Informationserhaltung dokumentiert. Die Annotatoren bearbeiten im Schnitt 10.000 Token pro Stunde. In einem weiteren Durchgang werden alle Unstimmigkeiten aufgelöst, die nicht automatisch beseitigt werden können. Der Vorgang wird als Adjudikation bezeichnet, an deren Ende eine erste Version des Goldstandards entsteht.

3.3. Zwischenergebnis: Häufigkeit von sensitiven Textstellen

Aus dem bereits annotierten und adjudizierten Teilkorpus lässt sich abschätzen, dass in den Urteilen ca. 35.000 sensitive Stellen pro 1 Million Token Fließtext zu finden sind. Das entspricht 70–80 Stellen pro Urteil durchschnittlicher Länge. Das häufigste Tag im Datensatz sind Datumsangaben: Im Schnitt finden sich ca. 20 Datumsangaben zu Sachverhalten und ca. 15 zur Prozessgeschichte pro Urteil (zusammen über 15.000 Vorkommen pro 1 Million Token); auf Weltwissen wird weit weniger als ein Mal pro Urteil verwiesen. In einem typischen Urteil findet man ca. 8 Adressangaben, zusätzlich ca. 10 identifizierende Merkmale von Adressen. Namen von natürlichen Personen werden im Schnitt ca. 10 Mal pro Urteil genannt (5.000 Vorkommen pro 1 Million Token), davon beziehen sich im Schnitt 4 Nennungen auf juristische Funktionsträger. Weitere identifizierende Merkmale für natürliche Personen kommen weniger als einmal pro Urteil vor. Juristische Personen werden seltener benannt (nur ca. 2–3 Mal pro Urteil) und hier kommen kaum weitere identifizierende Merkmale vor. Gerichte lassen sich im Durchschnitt 4 Mal pro Urteil und Aktenzeichen ca. 3 Mal pro Urteil finden.

4. Ergebnis

Ein entscheidender Weg zu komplexeren Legal-Tech-Anwendungen führt über die möglichst automatische Anonymisierung und damit Verfügbarmachung großer Mengen an Urteilen als notwendige Trainingsdaten für künftige subsymbolische KI-Anwendungen. Es liegt noch viel Forschungsarbeit zur juristischen Texterkennung, Auswertung und Anonymisierung vor uns, um feststellen zu können, ob eine automatische Anonymisierung von Urteilen technisch möglich ist und die Ergebnisse den rechtlichen Anforderungen genügen. Hierbei müssen stärker interdisziplinäre und empirische Ansätze verfolgt werden. Auch müssen die entscheidenden Rechtsfragen aufgearbeitet, gleichsam der Rechtsrahmen, der bei der Veröffentlichung und Anonymisierung von Entscheidungen eine Rolle spielt, noch stärker hinterfragt und durchdacht werden. Gerade wenn mit der Verfügbarmachung großer Mengen solcher Trainingsdaten eine Vielzahl an neuen Legal-Tech-Anwendungen entstehen und immense Potenziale für den Rechtsmarkt entfaltet werden sollten, skalieren die damit einhergehenden Risiken für Individuen. Daher ist Behutsamkeit gefragt, da eine unzureichende rechtliche Durchdringung der Problematik oder eine nicht zuverlässig evaluierte technische Umsetzung der automatischen Anonymisierung irreversible Folgen haben könnte, sobald diese Urteile in großer Zahl einmal veröffentlicht sind.

5. Literatur

ADRIAN, AXEL, Grundprobleme einer juristischen (gemeinschaftsrechtlichen) Methodenlehre, Duncker & Humblot, Berlin 2009.

ADRIAN, AXEL, Wie wissenschaftlich ist die Rechtswissenschaft?, RECHTSTHEORIE 2010, S. 521-548.
ADRIAN, AXEL, Grundzüge einer allgemeinen Wissenschaftstheorie auch für Juristen, Duncker & Humblot, Berlin 2014.
ADRIAN, AXEL, Der Richterautomat ist möglich – Semantik ist nur eine Illusion?, RECHTSTHEORIE 2017, S. 77-121.
ADRIAN, AXEL, Juristische Methodenlehre – Ein Vorbild für verantwortungsvolle Digitalisierung?. In: Schweighofer, Erich/Hötzendorfer, Walter/Kummer, Franz/Saarenpää, Ahti (Hrsg.), Verantwortungsbewusste Digitalisierung, Tagungsband des 23. Internationalen Rechtsinformatik Symposiums IRIS 2020, Bern 2020, S. 41-48.
ASHLEY, KEVIN, Artificial Intelligence and legal Analytics, Cambridge University Press, 2017.
BENCH-CAPON, TREVOR/SARTOR, GIOVANNI, A Model of Legal Reasoning with Cases Incorporating Theories and Values, Artificial Intelligence 150, 2003, S. 97-143.
BOEHME-NEßLER, VOLKER, Das Ende von der Anonymität, DuD 2016, S. 419-423.
BÜTTNER, WALTER/FROHN, MATTHIAS/SEEBACH, DANIEL, Elektronischer Rechtsverkehr und Informationstechnologie im Notariat, Beck, München 2019.
COUPETTE, CORINNA/FLECKNER, ANDREAS, Quantitative Rechtswissenschaft, JZ 2018, S. 379-389.
DÉVAUD, BLAISE/KUMMER, FRANZ, (Semi-)Automatische Anonymisierung von Entscheiden, Jusletter IT 23. Februar 2017.
FREUDENTHALER, BERNHARD, Case-based-Reasoning (CBR): Grundlagen und ausgewählte Anwendungsbeispiele des fallbasierten Schließens, AkademikerVerlag, Saarbrücken 2008.
FRIES, *MARTIN,* Rechtsdienstleistung durch KI. In: Kaulartz, Markus/Braegelmann, Tom (Hrsg.), Rechtshandbuch Artificial Intelligence und Machine Learning, Vahlen, München 2020, S. 651-661.
HAMMERBACHER, HANS, Die zu fordernde „ausreichende" Anonymisierung von Datensätzen, DuD 1984, S. 181-190.
HARTUNG, MARKUS, Gedanken zu Legal Tech und Digitalisierung. In: Hartung, Markus/Bues, Micha-Manuel/Halbleib, Gernot (Hrsg.), Legal Tech, Beck, München 2018, S. 5-17.
HORNUNG, GERRIT/WAGNER, BERND, Der schleichende Personenbezug, CR 2019, S. 565-574.
HORNUNG, GERRIT/WAGNER, BERND, Anonymisierung als datenschutzrelevante Verarbeitung?, ZD 2020, S. 223-228.
HAUPT, HEIKO, (Kein) Steuergeheimnis nach dem Finanzgerichtsprozess?, DStR 2014, S. 1025-1031.
KARG, MORITZ, Anonymität, Pseudonyme und Personenbezug revisited?, DuD 2015, S. 520-526.
KEUCHEN, MICHAEL, Grundbuch 4.0 – Folgt das Blockchain-Grundbuch dem Datenbankgrundbuch?, ZfIR 2020, S. 593-600.
KÖPFERL, GEORG, „Die Veröffentlichung unterbleibt im Hinblick auf das Steuergeheimnis", ZIS 2015, S. 375-385.
KUNTZ, WOLFGANG, Quantität gerichtlicher Entscheidungen als Qualitätskriterium juristischer Datenbanken, JurPC Web-Dok. 12/2006.
NÖHRE, INGO, Anonymisierung und Neutralisierung von veröffentlichungswürdigen Gerichtsentscheidungen, MDR 2019, S. 136-141.
PAAß, GERHARD/WAUSCHKUHN, UDO, Datenzugang, Datenschutz und Anonymisierung, R. Oldenbourg Verlag, München 1984.
POPPENHÄNGER, HOLGER, Die Übermittlung und Veröffentlichung statistischer Daten im Lichte des Rechts auf informationelle Selbstbestimmung, Duncker & Humblot, Berlin 1995.
PUTZKE, HOLM/ZENTHÖFER, JOCHEN, Der Anspruch auf Übermittlung von Abschriften strafgerichtlicher Entscheidungen, NJW 2015, S. 1777-1783.
ROCHER, LUC/HENDRICKX, JULIEN M./DE MONTJOYE, YVES-ALEXANDRE, Estimating the success of re-identifications in incomplete datasets using generative models, Nature Communications 10, 2019, Nr 3069.
SIMITIS, SPIROS/HORNUNG, GERRIT/SPIECKER, INDRA (Hrsg.), Datenschutzrecht, Nomos, Baden-Baden 2019.
SONNTAG, *MICHAEL*, Anonymisierung: Methoden und Zulässigkeit, Jusletter IT 25. Februar 2016.
VOGL, ROLAND, Changes in the US Legal Market Driven by Big Data/Predictive Analytics and Legal Platforms. In: Hartung/Bues/Halbleib (Hrsg.), Legal Tech, Beck, München 2018, S. 53-64.
VOKINGER, KERSTIN/MÜHLEMATTER, URS JAKOB, Re-Identifikation von Gerichtsurteilen durch «Linkage» von Daten(banken), Jusletter 2. September 2019.
WINTER, CHRISTIAN/BATTIS, VERENA /HALVANI, OREN, Herausforderungen für die Anonymisierung von Daten, ZD 2019, S.489-493.
WOLFF, AMADEUS/BRINK, STEFAN (Hrsg.), Beck`scher Online-Kommentar Datenschutzrecht, München, 34. Edition, Stand 01.11.2020.