

DISSEMINATION VON OPEN DATA



A Scoping Review on Analysis of the Barriers and Support Factors of Open Data
www.mdpi.com/2078-2489/15/1/5

ERFAHRUNGEN ZUR ANONYMISIERUNG UND WEITERVERWENDUNG VON PERSONENBEZOGENEN DATEN – QUALITATIVE ERGEBNISSE EINER EXPERTENBEFRAGUNG IM PROJEKT EASYANON

ERGEBNISSE

Hintergrund

Technologische Fortschritte und gesellschaftliche Implikationen von Open Data (OD)

→ OD zentrales Element gesellschaftlicher Transformation [1]: Technologische Möglichkeiten zur Erfassung/Analyse großer Datenmengen sind exponentiell gestiegen [1]

Herausforderungen im Datenschutz und der Nutzung anonymisierter personenbezogener Daten (pbD) als Open Data

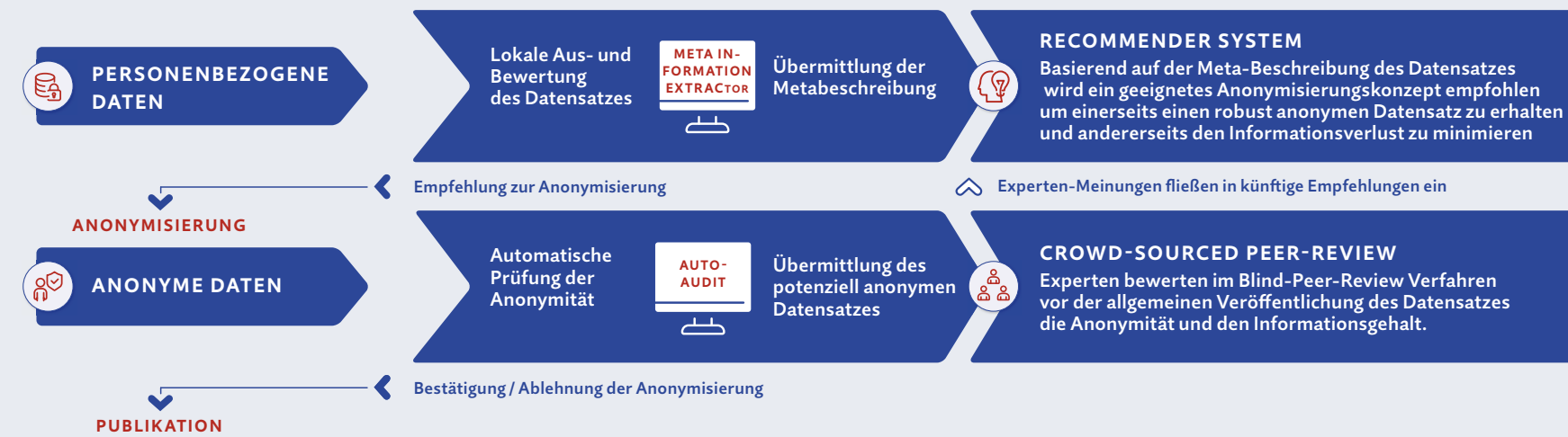
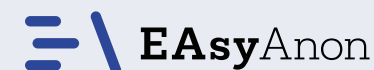
→ Balanceakt zwischen Datenschutzerfordernissen und dem maximalen Nutzen aus Daten [3]

Regulatorische Rahmenbedingungen und europäische Initiativen

→ DSGVO hat die Datenharmonisierung und -übertragbarkeit innerhalb Europas verbessert, Herausforderungen bei OD-Prozessen bestehen weiterhin [4]

Das Projekt EAsyAnon

Das Forschungsprojekt EAsyAnon entwickelt ein Softwaresystem zur effektiven Anonymisierung von personenbezogenen Daten.



Fragestellung

In der Stichprobe wurden Gesundheits-einrichtungen, Forschungseinrichtungen, Behörden und kleine und mittlere Unternehmen (KMU) befragt zu ...

- Arten und Formen von erhobenen pbD
- Erfahrungen zur Anonymisierung von pbD und deren anonymisierter Veröffentlichung als OD
- Barrieren und Förderfaktoren bezüglich Anonymisierung
- Unterstützungsleistungen
- Spezifischen rechtlichen und ethischen Herausforderungen

Methodik

- Ethikvotum
- Explorativ sequenzielles Mixed Methods Design
- Expert:inneninterviews
- Snowball-Sampling; 19 Interviews (Zeitraum Oktober 2023 – Januar 2024)
- Analyse: Strukturierende Inhaltsanalyse [5] mit MAXQDA

ARTEN UND FORMEN VON PERSONENBEZOGENEN DATEN

- Große Heterogenität von erhobenen pbD und Formen der Datenerhebung (meist schriftlich, wenig elektronisch)
- Viele unterschiedliche Datenformate, vor allem im Gesundheitssystem
- Datenstruktur sehr heterogen; teilweise sehr strukturierte Daten aufgrund von gesetzlichen Vorgaben oder elektronischer Erfassung
- Rohdaten oft unstrukturiert / handschriftlich, vorwiegend im Gesundheitsbereich
- Verwendungszwecke sind Forschungsgründe, interne Statistiken, zur Dokumentation, zum Support, zur Weiterentwicklung von Dienstleistungen, zur allgemeinen Abrechnung sowie aufgrund extern geforderter Bereitstellung
- Potenziell weitere Verwendungszwecke waren Weitergabe an Kunden, Register, Archive für Ausbildungszwecke, Qualitätssicherung, Werbemaßnahmen

„Ja wir erheben halt Daten zur Demografie, das ist Alter, das ist Geschlecht, das kann ganz viel sein [...] aus den Bewegungsdaten das sind also auch personenbezogene Daten, weil sie halt die Bewegung der Person abbilden“ (Transkription IP19, Pos. 32)

ERFAHRUNGEN ZU OPEN DATA UND ZUR ANONYMISIERUNG

- Heterogenes Verständnis zu OD; teilweise offene und kostenlose Zugänge, teilweise Zugangsbeschränkungen
- Homogenes Begriffsverständnis bei pbD. Hohe Sensibilität vorhanden, gerade bei Gesundheitsdaten.
- Stichprobe kaum Erfahrungen mit OD, außer im Forschungskontext, hier wurden anonymisierte pbD als OD veröffentlicht
- Wenn pbD veröffentlicht, dann meist in aggregierter Form
- Kaum Erfahrungen zur Anonymisierung, wenig durch Software, oft manuell anonymisiert, wenige interne Anonymisierungskonzepte vorhanden
- Hohes Interesse und Nachfrage an OD; hohe Bereitschaft zur Datenbereitstellung
- Eingebunden bei Anonymisierung: interne Fachkräfte (Datenschutz, Recht), externe Dienstleister
- Gründe für Nichtnutzung: keine Notwendigkeit, fehlende Eignung der Daten, drohende Konkurrenzsituationen, rechtliche Vereinbarungen

BARRIEREN UND FÖRDERFAKTOREN

- Barrieren zur Nutzung**
 - Wirtschaftliche Hürde (Kosten, Haftung, drohende Wettbewerbskonkurrenz, fehlender Innovationsschutz)
 - Persönliche Hürden (fehlendes Interesse/Vertrauen an Technik, mangelndes Engagement)
 - Datenaspekte (Sensibilität, Datenstruktur, kein Überblick zu eigenen Daten)
 - Technische Aspekte (fehlende Infrastruktur/IT-Fachkräfte)
 - Rechtliche Hindernisse (unklare/fehlende Rechtsgrundlagen zur Veröffentlichung/Zustimmungsverfahren, Föderalismus)
 - Ethische Hindernisse (Unumkehrbarkeit, mangelnde Kontrollfähigkeit, fehlender ethischer Bezugsrahmen, fehlende Transparenz zur Nutzung)
 - Staat und Behörden (schlechte Standortbedingungen, viel Bürokratie)

„In dem Bereich nervt der Föderalismus, weil jedes Bundesland hat seinen eigenen Datenschutzbeauftragten, hat sein eigenes Datenschutzgesetz, jetzt im Gesundheitswesen haben wir noch die Landeskrankenhausesetze“ (Transkription IP24, Pos. 77)

UNTERSTÜTZUNGSLEISTUNGEN

- Staatliche Unterstützungsleistungen (standardisierte Dokumente, Leitlinien, Best-Practice-Beispiele, finanzieller Unterstützung, Rechtsschutz für Haftung)
- Personelle Unterstützung (durch Fachkräften/Personalstellen/Abteilungen, externe Dienstleistende)
- technische Unterstützung (interoperable Software, im Idealfall Open-Source-Software)
- Förderfaktoren**
 - Wirtschaftliche Förderfaktoren (finanzielle Anreize, Unterstützung bei Veröffentlichung, vorhandene Ressourcen, kommerzielle Sicherheit, Kosteneinsparung)
 - Klare rechtliche Regeln (Haftung, Durchführung der Anonymisierung, Datenzugang)
 - Datenaspekte (Daten-Standards, hohe Datenqualität, Zugänglichkeit zu OD, Einsatz von Datentreuhändern, Transparenz in Verarbeitung/Outcome, etablierte Plattformen/Datenbanken)
 - Technische Aspekte (existierende Infrastruktur, Software mit hoher Usability)
 - Personelle Förderfaktoren (Interesse, Erfahrungen, Bereitschaft zur Nutzung)

ETHISCHE UND RECHTLICHE IMPLIKATIONEN

- Ethische Implikationen**
 - Daten als gesellschaftliches Machtinstrument (Teilhabe durch OD)
 - Gesellschaftlicher Konsens & Kostenfragen (finanzielle Belastungen durch Lizenzgebühren)
 - Zweckorientierung & Gemeinwohl (gesellschaftlicher Mehrwert, kein rein kommerzieller)
 - Herausforderungen & Akzeptanz (Vertrauen/Transparenz vs. Stigmatisierung/Missbrauch)
 - Regulierung & Zugangskontrolle (ethischer Kodex, klare Zugriffs- und Nutzungsbeschränkungen)
- Rechtliche Implikationen**
 - Datenschutz und rechtliche Rahmenbedingungen (zentrales Hemmnis, gesetzliche Konflikte bezüglich Datenlöschungsanspruch und Datenbereitstellungspflicht)
 - Regulatorische Anforderungen für KI-gestützte Anonymisierung (Berücksichtigung potenzieller Re-Identifizierungsrisiken, Klärung der Haftungsfragen im Kontext von Anonymisierungsverfahren)
 - Einwilligung und Transparenz (Einverständnis für Weiterverwendung als OD, Transparenz über OD-Zwecke, Wahrung der Selbstbestimmung und Eigenverantwortung)
 - Datenverwaltung und Eigentumsfragen (Datentreuhänder gewünscht, Klärung der Eigentumsverhältnisse von Rohdaten, Begrenzung der OD-Gültigkeitsdauer)
 - Technische Kontrolle und Regulierung (regelmäßige technische Kontrollen zur Anonymisierung, staatliche Regulierung/Zertifizierung durch Institutionen/Bundesamt für Sicherheit in der Informationstechnik (BSI), Unternehmen oder staatliche Stellen als potenzielle Prüfbehörden)

„Weil, ich denk immer, wenn man möchte, dass Medizin vorangetragen wird, muss man auch seinen Teil dazu beitragen, und wenn man nichts Besseres zu tun hat, als den ganzen Tag damit zu beschäftigt zu sein, dass man möglichst auf keinen Fall möchte, dass seine Daten nicht entschlüsselt werden kann, dann ist man falsch, weil das jedes auch anonymisierte Daten kann man auch mit den richtigen Tools entschlüsseln“ (Transkription IP24, Pos. 81)



EMPIRISCHE GRUNDLAGEN-UNTERSUCHUNG



DAS OUTCOME

Die Dissemination von Anonymisierungskonzepten und einer breiten Verwendung und Nutzung von Open Data hat bislang kaum stattgefunden. Oftmals stellen personelle Hürden, technische Barrieren und institutionelle Hindernisse eine große Herausforderung dar, welche den Erfahrungshorizont zu Anonymisierungen und Open Data aktuell begrenzt. Neben klaren rechtlichen und ethischen Handlungsrahmen wie unter anderem im Gesundheitsdatennutzungsgesetz (GDNG) formuliert, braucht es gezielte Unterstützungen, um eine breitere Nutzung von anonymisierten, pbD sind als OD zu ermöglichen.

Quellen [1] Deutscher Ethikrat (2017). Big Data und Gesundheit – Datensouveränität als informationelle Freiheitsgestaltung. Stellungnahme. Berlin. Online verfügbar unter <https://www.ethikrat.org/fileadmin/Publikationen/Stellungnahmen/deutsch/stellungnahme-big-data-und-gesundheit.pdf> [2] Viberg et al. (2022). What ethical approaches are used by scientists when sharing health data? An interview study. In: BMC medical ethics 23 (1), S. 41. DOI: 10.1186/s12910-022-00779-8. [3] Househ, Mowafa; Grainger; Rebecca; Petersen, Carolyn; Bamidis, Panagiotis; Merolli, Mark (2018). Balancing Between Privacy and Patient Needs for Health Information in the Age of Participatory Health and Social Media: A Scoping Review. In: Yearbook of medical informatics 27 (1), S. 29–36. DOI: 10.1055/s-0038-164197. [4] Feeney, Oliver; Werner-Felmayer, Gabriele; Siipi, Helena; Frischhut, Markus; Zullo, Silvia; Bartczko, Ursula et al. (2020). European Electronic Personal Health Records initiatives and vulnerable migrants: A need for greater ethical, legal and social safeguards. In: Developing world bioethics 20 (1), S. 27–37. DOI: 10.1111/dewb.12240. [5] Kuckartz, Udo (2018). Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung. 4., überarbeitete Auflage. Weinheim: Beltz Verlagsgesellschaft (Grundlagentexte Methoden). Online verfügbar unter <http://nbn-resolving.org/urn:nbn:de:bsz:31-eplflit-1138552>.

Autoren \ Projekt Konsortium

Norbert Lichtenauer¹, Johann Guggumus², Matthias Kampmann³, Julian Kis³, Florian Laumer⁴, Elena März², Florian Wahl⁵, Sebastian Wilhelm⁵

¹ Deggendorf Institute of Technology / Faculty Applied Health Sciences – Health Campus Bad Kötzing ² University of Augsburg / Faculty of Law ³ IT – SecurityCluster / Regensburg ⁴ PASSION4IT / Digital Transformation / Viechtach ⁵ Deggendorf Institute of Technology / Faculty Applied Computer Science – Technologie Campus Grafau **Contact: Norbert Lichtenauer / norbert.lichtenauer@th-deg.de**

