

OPEN PERSONAL DATA

ANONYMISIERUNG IM SPANNUNGSFELD ZWISCHEN INFORMATIONSGEHALT UND ROBUSTHEIT

MOTIVATION

- ↘ Daten bilden die Grundlage für Entscheidungen in Politik, Wissenschaft und Gesellschaft [1, 2]
- ↘ Wachsende Bedeutung von frei zugänglichen und lizenzfreien Daten (Open Data) [3]
- ↘ Herausforderung bei der Veröffentlichung von Daten als Open Data in Bezug auf Datenschutz, Sicherheit und Persönlichkeitsrechten [2, 4]

PROBLEMSTELLUNG

- ↘ Ohne explizite Rechtsgrundlage dürfen personenbezogene Daten gemäß DSGVO nicht veröffentlicht werden. Dies gilt aber nicht für anonyme Daten
→ Um personenbezogene Daten dennoch zu veröffentlichen, ist Anonymisierung der gängigste Ansatz [5]
- ↘ Konkrete Anweisungen wie eine Anonymisierung zu erfolgen hat, um den (rechtlichen) Bestimmungen zu genügen, fehlen jedoch [6]

SPANNUNGSFELD ZWISCHEN INFORMATIONSGEHALT UND ROBUSTHEIT DER ANONYMISIERUNG

- ↘ Anonymisierung von Daten erfordert häufig Definition von Zielparametern (z. B. k-Anonymisierung, l-Diversity)
- ↘ Parameterwahl bestimmt „Grad der Anonymisierung“
- ↘ Herausforderung: Auswahl robuster und rechtssicherer Parameter zur sicheren Anonymisierung
- ↘ Ziel: Verhindern von Identifizierbarkeit einerseits aber Erhalt des Informationsgehalts der Daten andererseits

Spannungsfeld zwischen zu wenig Anonymisierung (geringer Grad, identifizierbar) [7, 8, 9] und zu viel Anonymisierung (hoher Grad, Datenverlust an Nutzwert) [9].

| Name | Alter | Geschlecht | Arbeitsbereich | Wohnort (PLZ) | Wochenstunden | Einkommen |
|------|---------|------------|-------------------------------|---------------|---------------|-----------------|
| *** | 20 - 29 | m | Dienstleistungen und Handwerk | 5**** | 31 - 40 | 30.000 - 39.999 |
| *** | 30 - 39 | m | Dienstleistungen 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 |
| *** | 40 - 49 | m | Gesundheitswesen | 5**** | > 40 | 40.000 - 49.999 |

ANONYMISIERUNG MIT FOKUS AUF ROBUSTHEIT

⊕ HOHE ROBUSTHEIT ⊖ GERINGER INFORMATIONSGEHALT

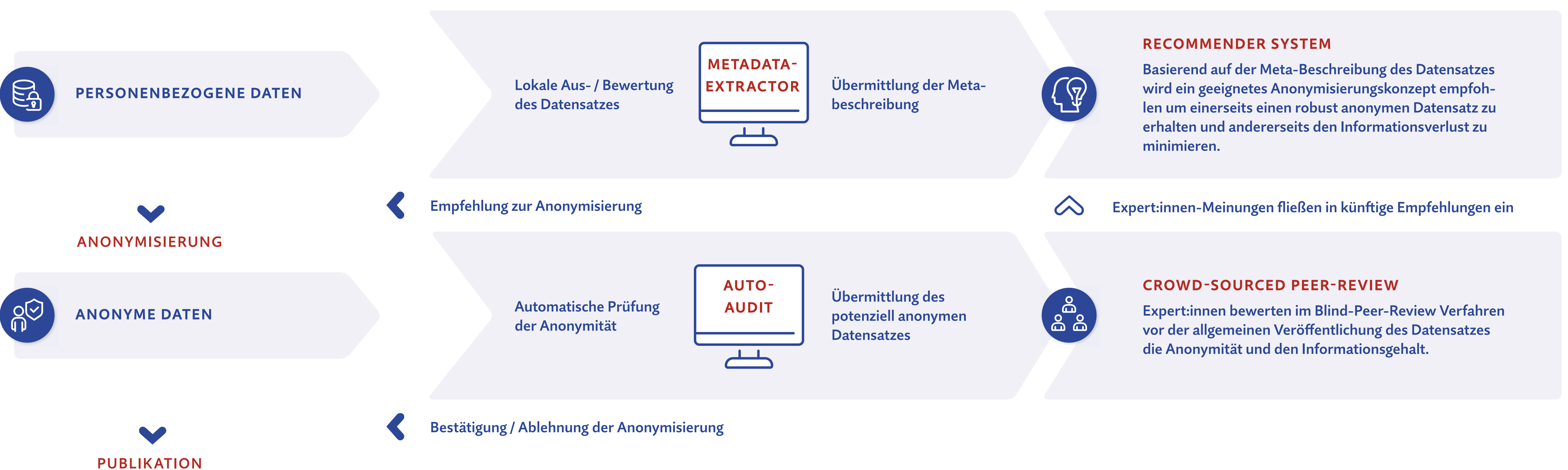
| Name | Alter | Geschlecht | Arbeitsbereich | Wohnort (PLZ) | Wochenstunden | Einkommen |
|-----------|-------|------------|----------------------|---------------|---------------|-----------|
| W. E**** | 24 | m | Postbote | 56284 | 39 | 34.100 |
| F. W***** | 30 | m | sebstst. Handwerker | 42113 | 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 |
| B. B***** | 18 | w | Auszubildende Köchin | 97289 | 38 | 10.200 |

ANONYMISIERUNG MIT FOKUS AUF INFORMATIONSGEHALT

⊖ GERINGE ROBUSTHEIT ⊕ HOHER INFORMATIONSGEHALT

ANSATZ EASYANON

- ↘ Erstellung von konkreten Empfehlungen zur Anonymisierung spezifischer Datensätze, die sowohl Robustheit als auch Informationsgehalt optimieren, unter Berücksichtigung rechtlicher und ethischer Anforderungen.
- ↘ Nachgelagerte Überprüfung der Anonymität eines Datensatzes mittels automatisierten und Expert:innen-basiertem Audit.



QUELLEN [1] European Commission, "Wertschöpfung durch Open Data," 2022. [Online]. [2] E. Tran and G. Scholtes, "Open Data Literature Review," Berkeley School of Law, University of California, 2015. [3] Bundesministerium des Innern und für Heimat, "Open Data," Oct. 2021. [Online]. [4] V. Demary, C. Askan, M. Fritsch, H. Goecke, T. Korte, A. Krotova, K. Lichtblau, and E. Schmitz, "Readiness data economy—bereitschaft der deutschen Unternehmen für die Teilhabe an der Datenwirtschaft," Institut der deutschen Wirtschaft Köln e. V., Köln, 2019. [5] J. Murray Jr, A. Mashadi, B. Lagesse, and M. Stiber, "Privacy Preserving Techniques Applied to CPNI Data: Analysis and Recommendations," arXiv preprint arXiv:2101.09834, 2021. [6] Kompetenzzentrum Open Data (CCOD), "Open Data Handbuch," Aug. 2020. [Online]. [7] L. Mirani, "London's bike-share program unwittingly revealed its cyclists' movements for the world to see," Apr. 2014. [Online]. [8] L. Sweeney, A. Abu, and J. Winn, "Identifying Participants in the Personal Genome Project by Name," arXiv preprint arXiv:1304.7605, 2013. [9] J. Manske, "Offene Daten und der Schutz der Privatsphäre," Oct. 2016. [Online].

PROJEKT EAsyAnon – Empfehlungs- und Auditsystem zur Anonymisierung

FÖRDERUNG BMBF **LAUFZEIT** 15.12.2022 – 14.12.2025 **PROJEKTLEITUNG** Prof. Dr. Florian Wahl, Technische Hochschule Deggendorf, florian.wahl@th-deg.de

KONTAKT Sebastian Wilhelm, sebastian.wilhelm@th-deg.de \ Jakob Folz, jakob.folz@th-deg.de \ Florian Wahl, florian.wahl@th-deg.de

PROJEKTPARTNER

