

# OPEN PERSONAL DATA

## ANONYMISATION IN THE CONFLICT BETWEEN ROBUSTNESS AND DATA UTILITY

### MOTIVATION

- Data forms the basis for decisions in politics, science, and society [1, 2]
- Growing importance of freely accessible and license-free data (open data) [3]
- Challenges in publishing data as open data in terms of privacy, security, and personal rights [2, 4]

### PROBLEM STATEMENT

- According to the GDPR, personal data may not be published without an explicit legal basis. However, this does not apply to anonymous data.
  - If personal data should be published, anonymisation is the most common approach [5]
- Concrete instructions on how to perform anonymisation to meet the (legal) requirements do not exist [6]

### CONFLICT BETWEEN DATA UTILITY AND ROBUSTNESS OF ANONYMISATION

- Anonymisation of data often requires definition of target parameters (e.g. k-anonymisation, l-diversity)
- Parameter selection determines the „degree of anonymisation“
- Challenge: Selection of robust and legally compliant parameters for secure anonymisation
- Goal: Mitigating the risk of re-identification on one hand whilst retaining the data's utility on the other

**Conflict between insufficient anonymisation (a low degree of anonymisation and possibly re-identifiable data has high utility value) [7, 8, 9] and too much anonymisation (a high degree of data utility loss) [9].**

Name	Age	Sex	Work	Residence (ZIP)	Hours/Week	Income
***	20 - 29	m	Services and crafts	5****	31 - 40	30.000 - 39.999
***	30 - 39	m	Services and crafts	4****	> 40	> 50.000
***	50 - 59	f	Office and administration	2****	31 - 40	> 50.000
***	30 - 39	f	Office and administration	5****	31 - 40	30.000 - 39.999
***	30 - 39	f	Art and culture	4****	21 - 30	< 19.999
***	40 - 49	m	Healthcare	5****	> 40	40.000 - 49.999

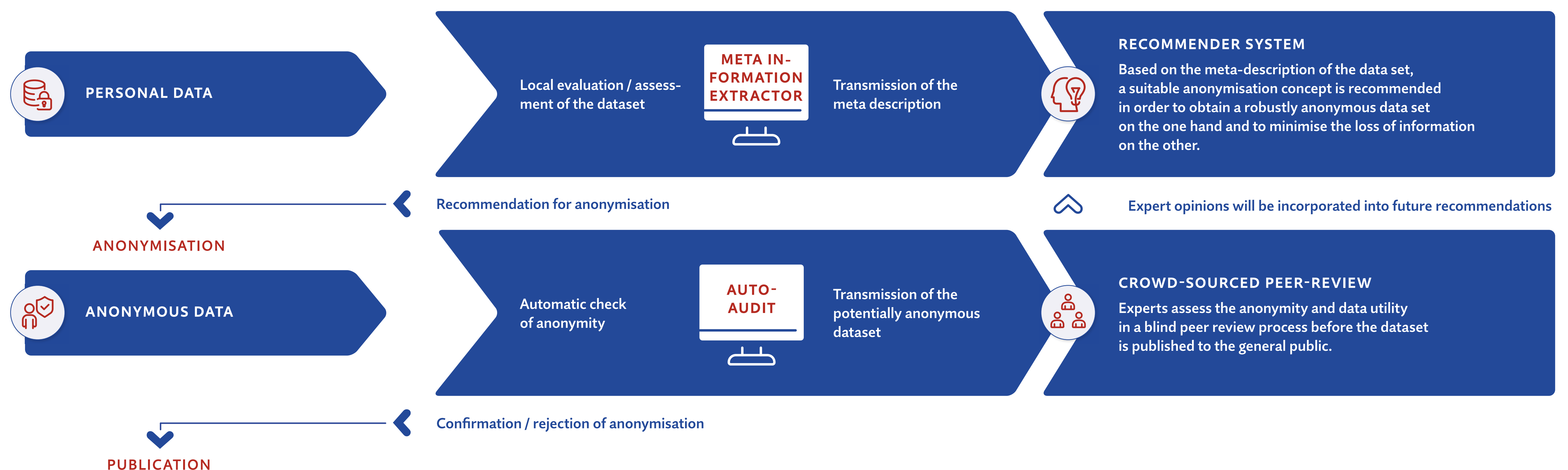
**ANONYMISATION EMPHASISING ROBUSTNESS**  
 + Strengthened robustness   - Reduced data utility

Name	Age	Sex	Work	Residence (ZIP)	Hours/Week	Income
W. E****	24	m	Letter carrier	56284	39	34.100
F. W*****	30	m	Self-employed craftsman	42113	50	64.300
J. S*****	56	f	Tax consultant	24626	40	79.540
S. B***	38	f	Office clerk	54450	39	38.500
K. H****	33	f	Musician	43221	20	19.800
G. R*****	40	m	Veterinarian	56307	45	49.000

**ANONYMISATION EMPHASISING DATA UTILITY**  
 - Reduced Robustness   + Preserved data utility

### EASYANON APPROACH

- Providing concrete recommendations for anonymising a specific dataset, optimising robustness and data utility, and considering legal and ethical requirements
- Conducting comprehensive audits of data set anonymity using an automated tool and expert assessments



SOURCE [1] European Commission, "Wertschöpfung durch Open Data," 2012. [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. Goetke, T. Korte, A. Krotova, K. Lichtblau, and E. Schmitt, "Readiness data economy - bereitchaft der deutschen Unternehmen für die Teilhabe an der Datenwirtschaft," Institut der deutschen Wirtschaft Köln e. V., Köln, 2019. [5] J. Murray Jr, A. Mashhad, B. Lagesse, and M. Stiber, "Privacy Preserving Techniques Applied to CPNI Data: Analysis and Recommendations," arXiv preprint arXiv:1001.09834, 2010. [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].

### PROJECT EAsyAnon – Empfehlungs- und Auditsystem zur Anonymisierung

**AUTHORS** Sebastian Wilhelm, Jakob Folz, Florian Wahl / Deggendorf Institute of Technology

**CONTACT** info-easyanon@th-deg.de

### PROJECT CONSORTIUM

