Session #6

Daten Fair teilen mit Differential Privacy: Die Möglichkeiten von Datenschutzgarantien für die Gesellschaft nutzbar machen

Dr. Daniel Franzen, FU Berlin

Prof. Dr. Claudia Müller-Birn, FU Berlin

freemove
Privacy-Centered Urban Mobility Data

GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

2016 IEEE 3rd International Conference on Data Science and Advanced Analytics (DSAA)

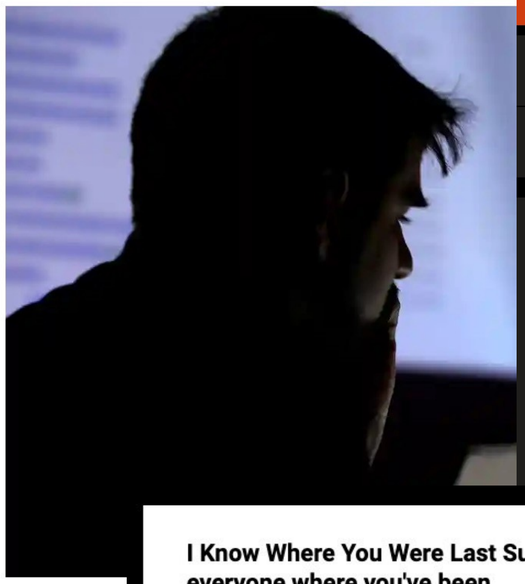
Anonymizing NYC Taxi Data: Does It Matter?

16, Pages: 140-148
okmark: 10.1109/DSAA.2016.21

's
ouriez
Doraiswamy
Freire

'Anonymous' browsing data can be easily exposed, researchers reveal

A journalist and a data scientist secured data from three million users easily by creating a fake marketing company, and were able to de-anonymise many users



'We wrote
clickstream f

A judge's

SPIEGEL Netzwelt Abonnement Anmelden

Menü | Startseite > Netzwelt > Web > Datenschutz-Debakel: Informatiker knacken anonymisierte Dat

Datenschutz-Debakel

Informatiker knacken anonymisierte Datenbank per Web-Suche

Der US-Filmverleiher Netflix veröffentlicht zu Forschungszwecken die Leihgeschichte seiner Kunden - natürlich ohne Namen. Doch
amen auf.
ioniert auch bei

I Know Where You Were Last Summer: London's public bike data is telling everyone where you've been

April 10, 2014

This article is about a publicly available dataset of bicycle journey data that contains enough information to track the movements of individual cyclists across London, for a six month period just over a year ago.

I'll also explore how this dataset could be linked with other datasets to identify the actual people who made each of these journeys, and the privacy concerns this kind of linking raises.

–

Menü | Süddeutsche Zeitung

Meine SZ | SZ Plus | Ukraine | Oktoberfest | Politik | Wirtschaft | Meinung | Panorama | Sport | München | Kultur | Medien
ITATION

Metadaten

Was die Kreditkarte verrät

30. Januar 2015, 10:46 Uhr

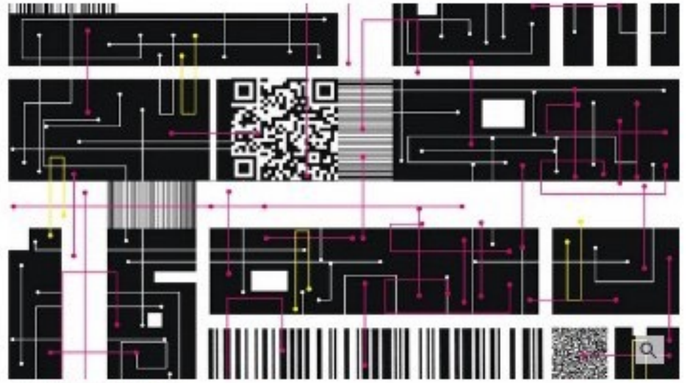


Illustration: Stefan Dimitrov

- Selbst große anonymisierte Datensätze bieten unter Umständen nur wenig Schutz, berichten Forscher im Fachmagazin *Science*.
- Die Wissenschaftler analysierten "einfach anonymisierte" Kreditkartentransaktionen von 1,1 Millionen Menschen über einen Zeitraum von drei Monaten hinweg.
- Anhand weniger Anhaltspunkte - wo jemand seinen Kaffee trank oder ins Restaurant ging etwa - konnten die Forscher die meisten Personen in der Datenbank reidentifizieren.

Differential Privacy als Lösung?

Ziel

Privatsphäre von Datensätzen bei statistischen Abfragen

Garantie

Daten einzelner keinen nennenswerten Einfluss auf das Ergebnis von Abfragen

Maß

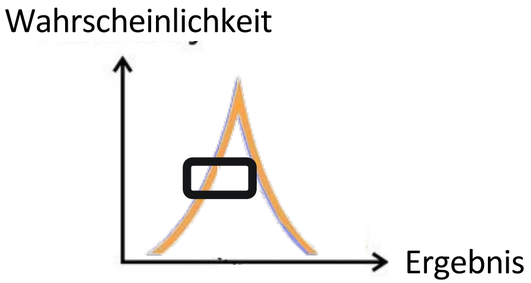
Privacy Budget ε (Epsilon) begrenzt den Einfluss

Mathematische Definition

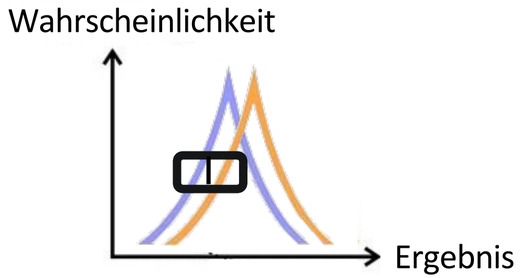
$$\frac{P\left[f\left(D1\right) \in S\right]}{P\left[f\left(D2\right) \in S\right]} \leq e^{\varepsilon}$$

GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

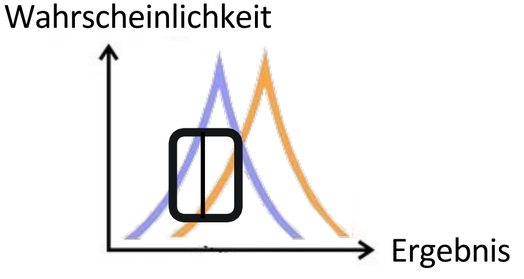
Kompromiss – Privatsphäre / Aussagekraft?



Mit $\varepsilon = 0$



Mit $\varepsilon = 1$



Mit $\varepsilon = 8$

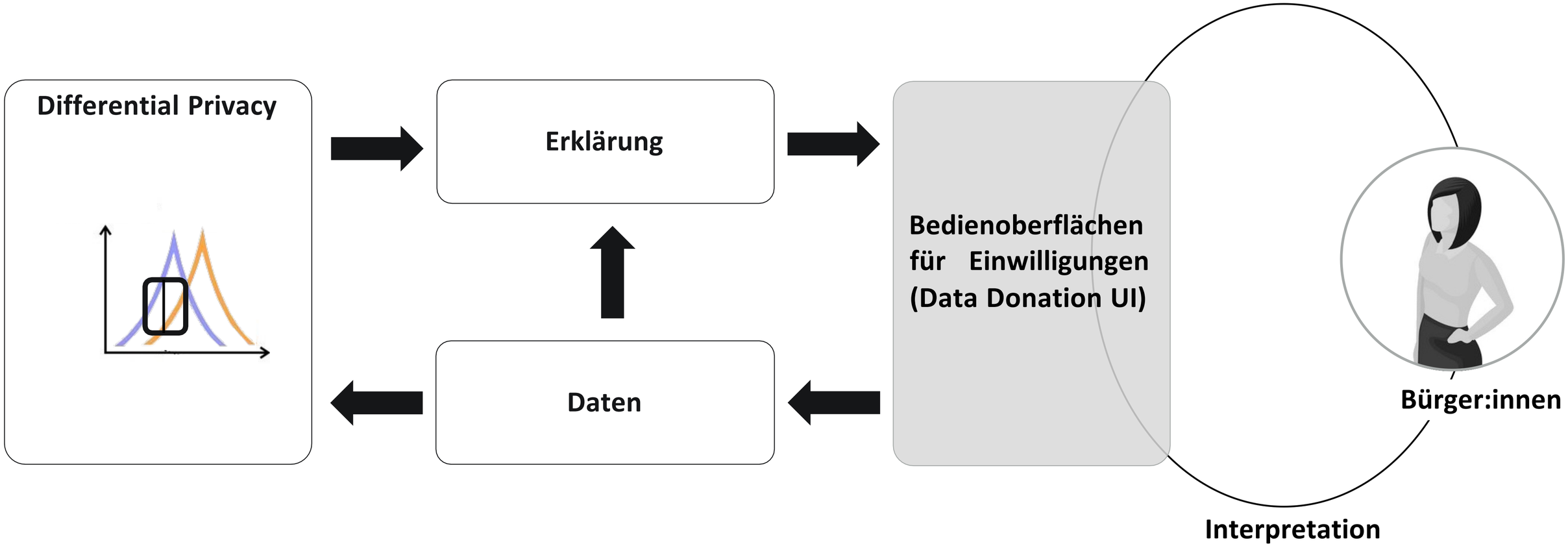
Privatsphäre ↑ ↘ ↓

Aussagekraft ↓ ↗ ↑

freemove
Privacy-Centered Urban Mobility Data

GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

Erklärbarkeit von Datenschutzmechanismen



Sörries, Peter; Müller-Birn, Claudia; et al.(2021): Privacy Needs Reflection: Conceptional Design Rationales for Privacy-Preserving Explanation User Interfaces. Mensch und Computer 2021 - Workshopband. DOI: 10.18420/muc2021-mci-wsc-389. Bonn: Gesellschaft für Informatik e.V.. MCI-WS14: Usable Security und Privacy Workshop.

Herausforderungen in der Forschung

Wie können wir dazu beitragen, einen gesellschaftlichen Aushandlungsprozess in Bezug auf Privatsphäre zu fördern?

Der Mehrwert von
DP entsteht durch
die Wahl privacy Budgets ε

Möglichkeiten der Kommunikation
des Privatsphäreverlusts sind
bisher begrenzt

freemove
Privacy-Centered Urban Mobility Data

GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

Herausforderungen in der Forschung

Wie können wir dazu beitragen, einen gesellschaftlichen Aushandlungsprozess in Bezug auf Privatsphäre zu fördern?

Der Mehrwert von DP entsteht durch die Wahl privacy Budgets ε

Möglichkeiten der Kommunikation des Privatsphäreverlusts sind bisher begrenzt

Risiko für die Privatsphäre wird bisher unzureichend kommuniziert

Risikokommunikation zum Schutz der Privatsphäre noch nicht gut erforscht

freemove
Privacy-Centered Urban Mobility Data

GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

Herausforderungen in der Forschung

Wie können wir dazu beitragen, einen gesellschaftlichen Aushandlungsprozess in Bezug auf Privatsphäre zu fördern?

Der Mehrwert von DP entsteht durch die Wahl privacy Budgets ε

Möglichkeiten der Kommunikation des Privatsphäreverlusts sind bisher begrenzt

Risiko für die Privatsphäre wird bisher unzureichend kommuniziert

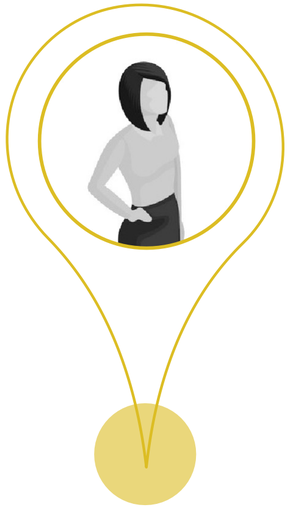
Risikokommunikation zum Schutz der Privatsphäre noch nicht gut erforscht

freemove
Privacy-Centered Urban Mobility Data

GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

Privacy Budget $\varepsilon \longrightarrow$ Risiko

Wahre Teilnahme



Zufall (n=2) [2]

$$\text{Risiko} = \frac{1}{1 + e^{-\varepsilon}}$$

Luise Mehner, Saskia Nuñez von Voigt, and Florian Tschorsch. 2021. Towards Explaining Epsilon: A Worst-Case Study of Differential Privacy Risks. In 2021 IEEE European Symposium on Security and Privacy Workshops (EuroS PW). IEEE, Vienna, Austria, 328–331. https://doi.org/10.1109/EuroSPW54576.2021.00041

Herausforderungen in der Forschung

Wie können wir dazu beitragen, einen gesellschaftlichen Aushandlungsprozess in Bezug auf Privatsphäre zu fördern?

Der Mehrwert von DP entsteht durch die Wahl privacy Budgets ε

Möglichkeiten der Kommunikation des Privatsphäreverlusts sind bisher begrenzt

Risiko für die Privatsphäre wird bisher unzureichend kommuniziert

Risikokommunikation zum Schutz der Privatsphäre noch nicht gut erforscht

GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

Studie 1 – Textuelle Risikoformate

Verwendete Risikoformate

Prozentsatz

Das Ereignis tritt in 25 % der Fälle ein.

Häufigkeit (Frequenz)

Das Ereignis tritt in 15 von 60 Fällen ein.

Erweiterung

Ohne Ergänzung

Positives / Negatives Framing

Das Ereignis tritt
in 75 % der Fälle / in 45 von 60 Fällen
NICHT ein.

Vergleich zu Status Quo

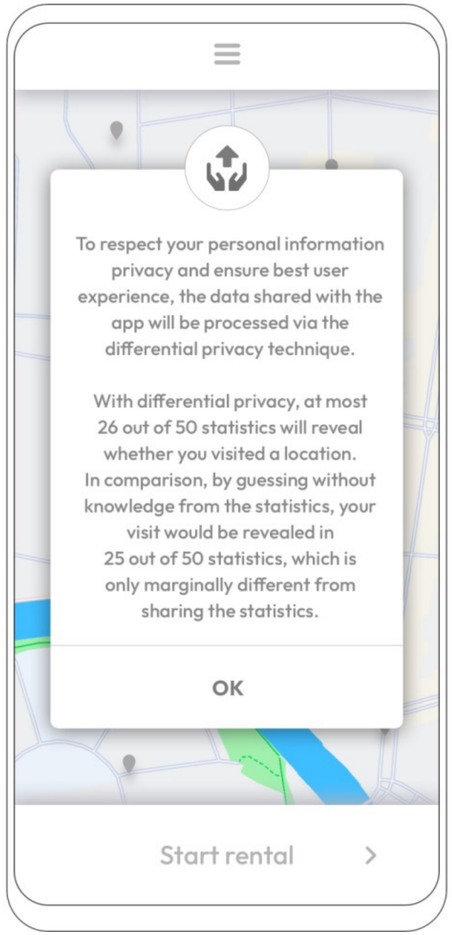
Ohne Ihre Teilnahme tritt das Ereignis
in 20 % der Fälle / in 12 von 60 Fällen
ein.

Daniel Franzen, Saskia Nuñez von Voigt, Peter Sörries, Florian Tschorsch, and Claudia Müller-Birn. 2022. Am I Private and If So, how Many? Communicating Privacy Guarantees of Differential Privacy with Risk Communication Formats. In Proc. ACM SIGSAC Conference on Computer and Communications Security (CCS '22). 1125–1139. https://doi.org/10.1145/3548606.3560693

freemove
Privacy-Centered Urban Mobility Data

GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

Studie 1 – Studienerkenntnisse



Privatsphäre-Risiken sind grundsätzlich für Datenschutzentscheidungen geeignet

Bedienoberflächen müssen noch stärker auf Risiko abgestimmt werden (Studie 2)

Subjektives Verständnis von Privatsphäre-Risiko muss stärker unterstützt werden (Studie 2)

Falsches Sicherheitsgefühl gefährdet Privatsphäre von vulnerablen Gruppen

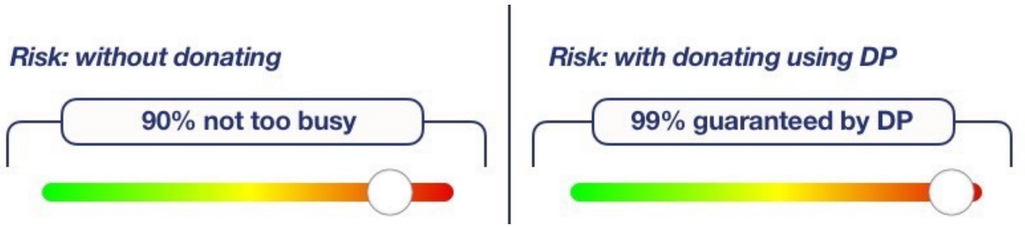
Statistisches Wissen muss bei der Gestaltung berücksichtigt werden

Daniel Franzen, Saskia Nuñez von Voigt, Peter Sörries, Florian Tschorsch, and Claudia Müller-Birn. 2022. Am I Private and If So, how Many? Communicating Privacy Guarantees of Differential Privacy with Risk Communication Formats. In Proc. ACM SIGSAC Conference on Computer and Communications Security (CCS '22). 1125–1139. https://doi.org/10.1145/3548606.3560693

Studie 2 – Interaktiv-Visuelle Risikoformate

“[I did not share because] The probability percentage was kind of high - over 50%”

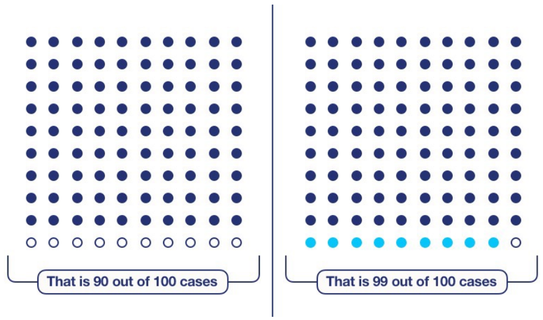
Das Privatsphäre-Risiko hängt von individuellen Faktoren (z.B. Vor-Risiko) ab



Interaktive Exploration hilft das Privatsphäre-Risiko in der individuellen Situtation zu evaluieren

“[I am missing] more details about what the 53% number means?”

Numerische Risiken sind schwer vorzustellen



Visualisierungen helfen beim Verständnis

Daniel Franzen, Claudia Müller-Birn and Odette Wegwarth 2023. In Proc. of the ACM on Human-Computer Interaction, Issue CSCW (accepted for publication)

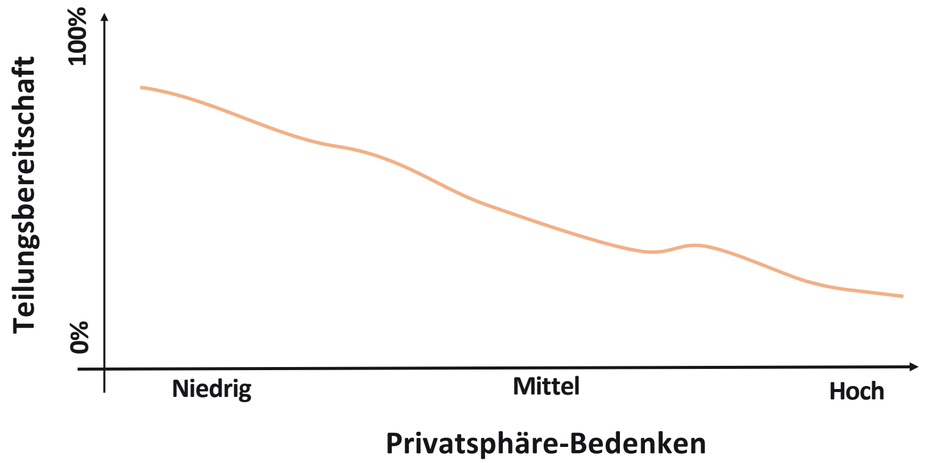
Studie 2 – Informierte Entscheidung

> [Informed decision is a] choice [...] made [...] **using relevant information** about the advantages and disadvantages of all the possible courses of action, **in accord with the individual's beliefs**
>
> Hilary Bekker, J. G. Thornton, C. M. Airey, JlBl Connelly, J. Hewison, M. B. Robinson, J. Lilleyman, M. MacIntosh, A. J. Maule, and S. Michie. 1999. Informed decision making: an annotated bibliography and systematic review. Health Technol Assess 3, 1 (1999), 1–156.

Wahl:
Datenspende mit DP Ja / Nein

Relevante Informationen:
ε / Risikoinformation

Individuelle Überzeugung:
generelle Privatsphäre-Bedenken



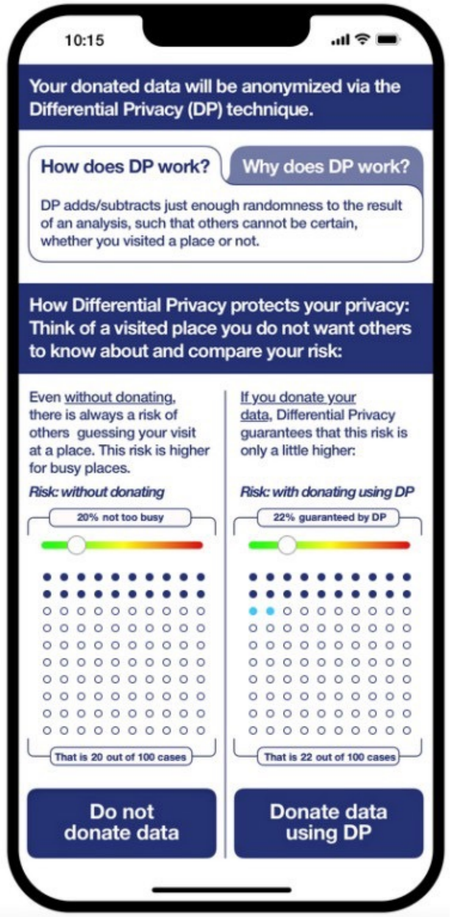
Daniel Franzen, Claudia Müller-Birn and Odette Wegwarth 2023. In Proc. of the ACM on Human-Computer Interaction, Issue CSCW (accepted for publication)

Studie 2 – Studienaufbau

Allgemeine DP Information

Risiko-Format

- **Text**
- **Interaktive Exploration**
- **Visualisierung**



Wie kann das Verständnis des Risikos durch Darstellungselemente unterstützt werden?

Studienkontext

- Car-sharing App
- Entscheidung für / gegen Datenspende

Studienaufbau

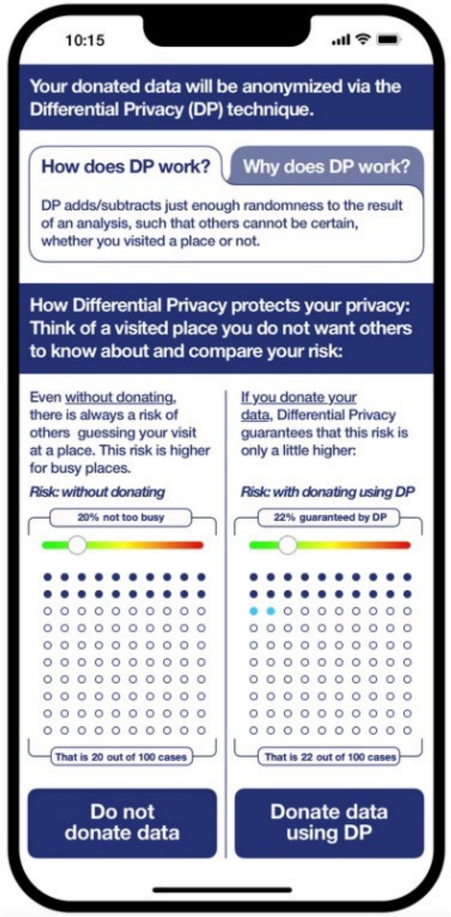
- Online-Experiment auf MTurk mit 304 Teilnehmenden
- Interaktiver Click-Prototyp mit 5 unterschiedlichen UI-Design (~ 75 pro Gruppe)
- Fragebögen zur Erfassung der Usability, der Privatsphäre-Einstellung, der Einprägsamkeit des Risikos / der Entscheidung sowie statistisches Vorwissen, Entscheidungstyp (GDMS), Kognitionsbedürfnis (NFC)

Daniel Franzen, Claudia Müller-Birn and Odette Wegwarth 2023. In Proc. of the ACM on Human-Computer Interaction, Issue CSCW (accepted for publication)

freemove
Privacy-Centered Urban Mobility Data

GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

Studie 2 – Studienerkenntnisse



- Weder Interaktion noch Visualisierung allein fördern eine informierte Entscheidung, aber Visualisierung mit → Die Art der Darstellung scheint das mentale Modell der Nutzenden zu unterstützen.
- Daten-Teilungsverhalten im Ganzen wird nicht durch die bereitgestellten Erklärungen beeinflusst (weder mehr Teilungen noch weniger) → Der Vorbehalt, dass Risikoinformation eher zu weniger Datenteilungen führt ist nicht berechtigt.
- Bedienoberflächen mit quantitativen Informationen werden nicht als weniger bedienfreundlich eingestuft.
- Statistisches Wissen beeinflusst die Einprägsamkeit des Risikos.

Daniel Franzen, Claudia Müller-Birn and Odette Wegwarth 2023. In Proc. of the ACM on Human-Computer Interaction, Issue CSCW (accepted for publication)

Fazit

Differential Privacy kann es der Gesellschaft ermöglichen, von Big Data zu profitieren und gleichzeitig die individuelle und kollektive Privatsphäre zu schützen.

freemove
Privacy-Centered Urban Mobility Data

GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

Fazit

Differential Privacy kann es der Gesellschaft ermöglichen, von Big Data zu profitieren und gleichzeitig die individuelle und kollektive Privatsphäre zu schützen.

Die Akzeptanz für Datenspenden kann nur durch einen transparenten und verständlichen PET-Einsatz über die Risiko-Kommunikation sichergestellt werden.

freemove
Privacy-Centered Urban Mobility Data

GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

Fazit

Differential Privacy kann es der Gesellschaft ermöglichen, von Big Data zu profitieren und gleichzeitig die individuelle und kollektive Privatsphäre zu schützen.

Die Akzeptanz für Datenspenden kann nur durch einen transparenten und verständlichen PET-Einsatz über die Risiko-Kommunikation sichergestellt werden.

Partizipative und werte-orientierte Gestaltungsansätze helfen dabei, die sozialen Folgen von technologischen Innovationen offenzulegen.

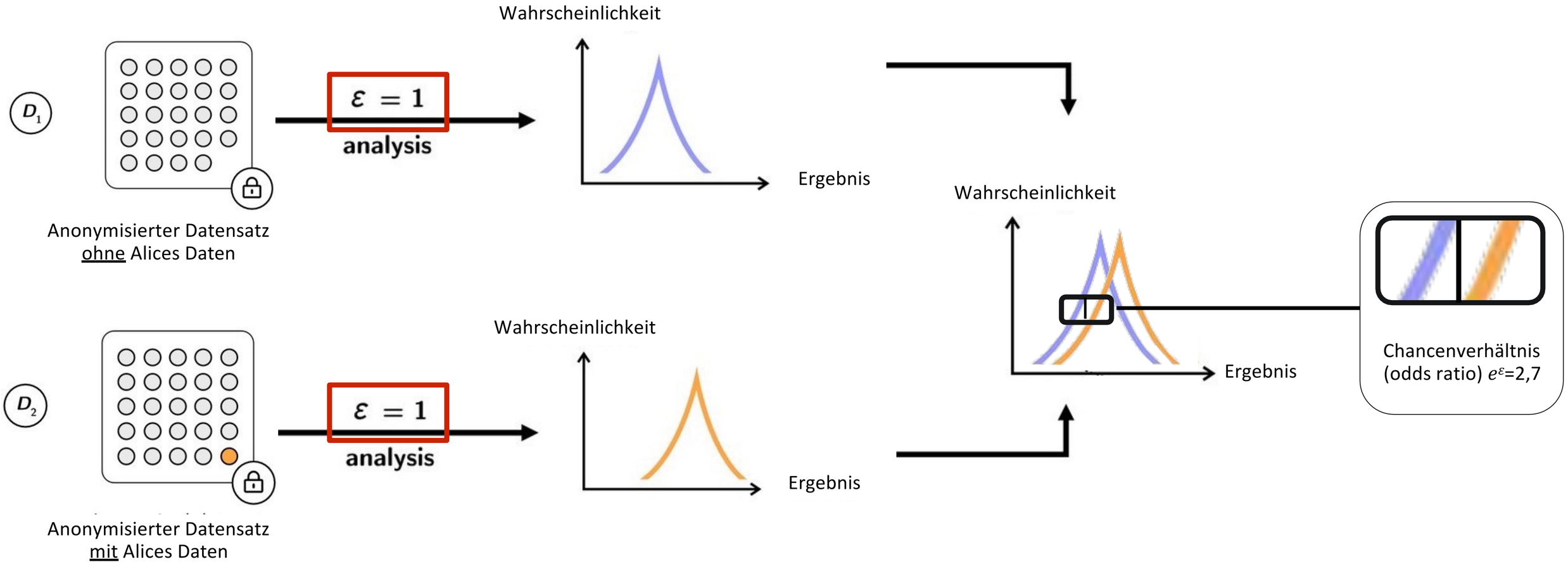
freemove
Privacy-Centered Urban Mobility Data

GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

Vielen Dank

GEFÖRDERT VOM
Bundesministerium
für Bildung
und Forschung

Differential Privacy – illustriert

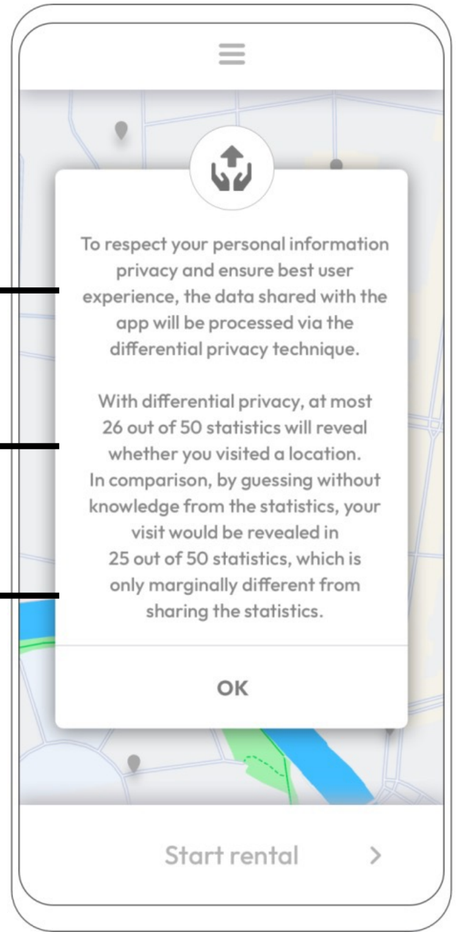


Studie 1 – Studienaufbau

Allgemeine DP Information

Risiko-Format

Ergänzung / Variation



Welches Risikoformat kann Privatsphäre Risiko verständlich darstellen?

Studienkontext

- Fiktive Car-Sharing App, die Daten zu Fahrthistorien anonymisiert der Stadt zur Verfügung stellen will.

Studienaufbau

- Online-Experiment auf MTurk mit 343 Teilnehmenden
- Interaktiver Click-Prototyp in 7 Varianten (~ 45 Teilnehmende pro Gruppe)
- Kontrollgruppe ohne quantitative Risikoformate
- Fragebögen zur Erfassung objektives und subjektives Verständnis sowie Statistisches Wissen, Privatsphäre-Einstellung, Geschlecht

Daniel Franzen, Saskia Nuñez von Voigt, Peter Sörries, Florian Tschorsch, and Claudia Müller-Birn. 2022. Am I Private and If So, how Many? Communicating Privacy Guarantees of Differential Privacy with Risk Communication Formats. In Proc. ACM SIGSAC Conference on Computer and Communications Security (CCS '22). 1125–1139. https://doi.org/10.1145/3548606.3560693

Herausforderungen in der Forschung

Wie können wir dazu beitragen, einen gesellschaftlichen Aushandlungsprozess in Bezug auf Privatsphäre zu fördern?

?

Der Mehrwert von DP entsteht durch die Wahl privacy Budgets ε

↓

Möglichkeiten der Kommunikation des Privatsphäreverlusts sind bisher begrenzt

Risiko für die Privatsphäre wird bisher unzureichend kommuniziert

↓

Risikokommunikation zum Schutz der Privatsphäre noch nicht gut erforscht

Die Bedeutung der Privatsphäre hängt von den Werten ab

↓

Ansätze zur systematischen Erhebung von Werten noch wenig entwickelt

freemove
Privacy-Centered Urban Mobility Data

GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

Studie 3 – Studienaufbau

Selection of Workshop Outcomes (Image Credits: Sörries, P.)

Studienkontext

- Workshop zur partizipativen Erhebung von Werten im Zusammenhang mit der Spende von Mobilitätsdaten

Studienaufbau

- Drei Workshops (3/4 2023) mit 13 Teilnehmer:innen unterschiedlichem Hintergrunds (Studierende, Forschende, Personen aus der Stadtplanung oder Digitalisierung)

Analyse der Daten

- Anwendung eines induktiven Kodierverfahrens zur Ermittlung der Werte und Wertekonflikte
- Anwendung eines deduktiven Kodierverdahrens zur Analyse der Werte-Szenarien basierend auf den Werten

Peter Sörries, Daniel Franzen, Markus Sperl, and Claudia Müller-Birn. 2023. Foregrounding Values through Public Participation: Eliciting Values of Citizens in the Context of Mobility Data Donation. In Proceedings of Mensch und Computer 2023 (MuC '23). ACM, 387–394. https://doi.org/10.1145/3603555.3608531

freemove
Privacy-Centered Urban Mobility Data

GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

Studie 3 – Partizipativer Workshop zur Werteermittlung

Exploration

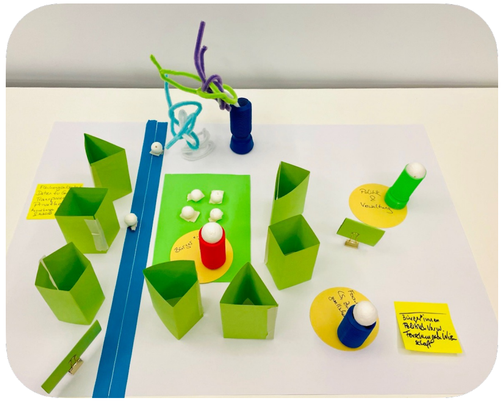
Die Teilnehmenden erkunden ihre Werte im Hinblick auf den Workshop-Kontext.

Systematisierung

Die Teilnehmenden verorten aus einer Wertekarte ihre Werte aus Phase 1 in Anbetracht direkter und indirekter Interessengruppen, um mögliche Wertekonflikte zu adressieren.

Übersetzung

Auf Grundlage der Wertekarte gestalten die Teilnehmenden Werteszenarien, um eine ideale Situation zu illustrieren.



Reflexion

Die Teilnehmenden reflektieren über ihre die Ergebnisse und Erfahrungen der einzelnen Workshopphasen.

Peter Sörries, Daniel Franzen, Markus Sperl, and Claudia Müller-Birn. 2023. Foregrounding Values through Public Participation: Eliciting Values of Citizens in the Context of Mobility Data Donation. In Proceedings of Mensch und Computer 2023 (MuC '23). ACM, 387–394. https://doi.org/10.1145/3603555.3608531

Studie 3 – Studienerkenntnisse

Konzentration auf lokale Infrastrukturen zur Stärkung der Selbstverwaltung der Bürger:innen (z. B. Datenerfassung in bestimmten Lebensbereichen zur Förderung einer engen Kommunikation zwischen den Mitgliedern der Gemeinschaft).

Betonung der sozialen Interaktion, um die Reflexion der Bürger:innen über die Datenspende zu fördern (z. B. kollektiv genutzte Technologien).

Einrichtung von Datenpraktiken, um die Daten-souveränität der Bürger zu stärken (z. B. durch die Einrichtung eines Datenvermittlers für Datenver-waltung auf Forschungsdatenplattformen).

Peter Sörries, Daniel Franzen, Markus Sperl, and Claudia Müller-Birn. 2023. Foregrounding Values through Public Participation: Eliciting Values of Citizens in the Context of Mobility Data Donation. In Proceedings of Mensch und Computer 2023 (MuC '23). ACM, 387–394. https://doi.org/10.1145/3603555.3608531

Session #6: Daten Fair teilen mit Differential Privacy: Die Möglichkeiten von Datenschutzgarantien für die Gesellschaft nutzbar machen

freemove
Privacy-Centered Urban Mobility Data

GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung