

Datenanonymisierungs- plattform (DAP)

Praktischer Lösungsansatz für TK-Daten, DatenTag, 07.12.2022

Timo Wilken, CIPP/E, CIPM, CIPT, Data Protection & Regulation Office

Herausforderung.

Daten anonymisieren und dabei den Wert der Informationen erhalten.

01

Strikte Vorgaben zur Verarbeitung von TK-Daten.

- Bei o2 Telefónica fallen täglich mehrere Milliarden Events bei der Erbringung von TK-Diensten an.
- Der Wert dieser Informationen kann auf personenbezogener Ebene jedoch nicht ausgeschöpft werden, da die möglichen Verarbeitungen von TK-Daten gesetzlich sehr eingeschränkt sind – und sogar strikter reguliert als die Verarbeitung „normaler“ personenbezogener Daten.
- Unser Lösungsansatz: Anonymisierung – dabei war und ist es aber wichtig, den Wert der Informationen möglichst hoch zu halten, um den höchstmöglichen Nutzen aus ihnen ziehen zu können.
- In enger Abstimmung mit dem BfDI hat o2 Telefónica daher bereits vor einigen Jahren eine Daten-anonymisierungsplattform (DAP) für die Anonymisierung von Bestands- und Standortdaten entwickelt.



Praktische Fragen: Relativer Anonymitätsbegriff und Unverhältnismäßigkeit.

§ 3 Abs. 6 BDSG a.F.

„**Anonymisieren** ist das Verändern personenbezogener Daten derart, dass die Einzelangaben über persönliche oder sachliche Verhältnisse nicht mehr oder nur mit einem **unverhältnismäßig hohem Aufwand an Zeit, Kosten und Arbeitskraft** einer bestimmten oder bestimmbaren natürlichen Person zugeordnet werden können.“

Unverhältnismäßigkeit als zentrale Bedingung im Zusammenspiel mit dem potentiellen Erkenntnisgewinn.

ErwGr 26 zur DSGVO

„Um festzustellen, ob eine natürliche Person identifizierbar ist, sollten alle Mittel berücksichtigt werden, die von dem Verantwortlichen oder einer anderen Person **nach allgemeinem Ermessen wahrscheinlich genutzt werden**, um die natürliche Person direkt oder indirekt zu identifizieren, wie beispielsweise das Aussondern. Bei der Feststellung, ob Mittel nach allgemeinem Ermessen wahrscheinlich zur Identifizierung der natürlichen Person genutzt werden, sollten **alle objektiven Faktoren, wie die Kosten der Identifizierung und der dafür erforderliche Zeitaufwand**, herangezogen werden, wobei die **zum Zeitpunkt der Verarbeitung verfügbare Technologie und technologische Entwicklungen** zu berücksichtigen sind.“

Praktische Fragen: Ausschluss möglicher Re-Identifikationsansätze.

Direkter Personenbezug

Mögliche Re-Identifikation durch einzelne Eigenschaften, die eine Person **direkt und eindeutig** identifiziert.

Indirekter Personenbezug

Mögliche Re-Identifikation durch einzelne Eigenschaften, die sich mit entsprechendem **Zusatzwissen** einer Person eindeutig zuordnen lassen.

Dynamisches Profil.

Mögliche Re-Identifikation durch eine Kombination einer Vielzahl dynamisch anfallender Daten, so dass das entstehende Muster mit entsprechendem **Zusatzwissen** einen eindeutigen Rückschluss auf eine Person zulässt.

Statischer Fingerabdruck.

Mögliche Re-Identifikation durch Nutzung einer, aus **mehreren statischen Eigenschaften zusammengesetzten**, eindeutigen statischen Kennung für die Erstellung langfristiger dynamischer Profile oder eines direkten / indirekten Personenbezugs.

Lösung.

Mehrstufiges Anonymisierungs-verfahren
in verschiedenen Zonen.

02

Mehrstufiges Anonymisierungsverfahren: Vereinfachte Darstellung.

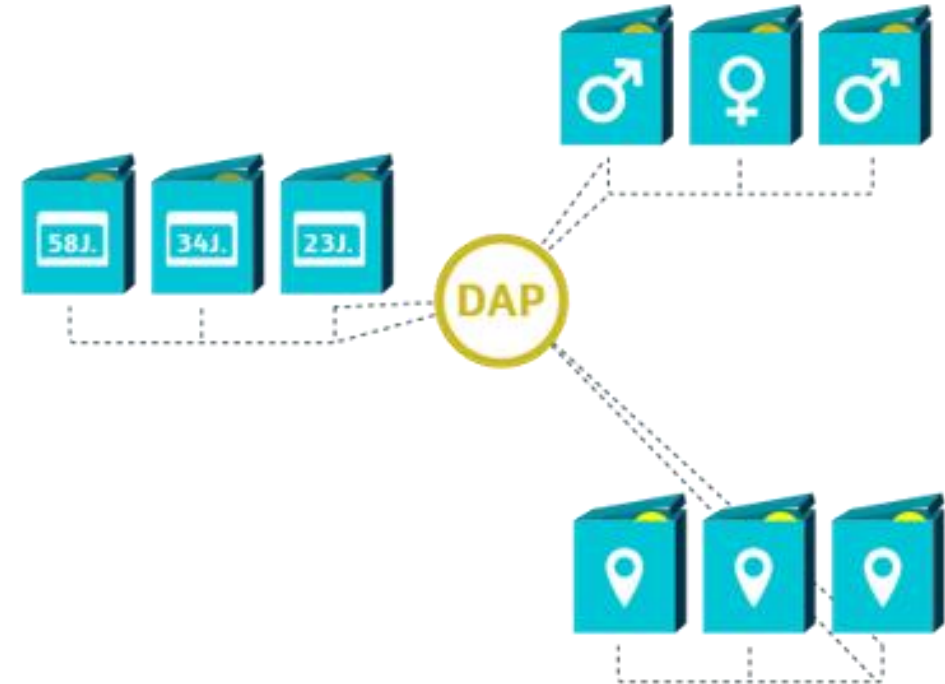
Vereinfacht dargestellt, besteht das Anonymisierungsverfahren darin, einzelne „Bewegungsspuren“ über einen jeweils begrenzten Zeitraum zu erzeugen, die keiner Person mehr zugeordnet oder miteinander verknüpft werden können.

- Dabei wird zunächst aus der **IMSI** (International Mobile Subscriber Identity) **und einem täglich wechselnden „Salt“** eine **neue Kennung** berechnet. Ein **Zurückrechnen** ist damit **nicht mehr möglich**. Wegen des wechselnden „Salt“ können die **Bewegungsspuren nicht** über **mehrere Tage** miteinander **verknüpft** werden.
- Dieser Hashwert stellt eine Kennung dar, die für **ein Jahr** einer Mobilfunkkarte zugeordnet ist; eine Rückrechnung auf die IMSI ist nicht möglich. Diese Jahreskennung wird zusammen mit einem für **einen Tag gültigen Salt** mit einem kryptographischen Schlüssel verschlüsselt. Das Ergebnis ist **eine für einen Tag gültige Kennung für eine Mobilfunkkarte**.
- **Bestandsdaten** über das Alter, Geschlecht und Wohnort (auf – teilweise gekürzter – PLZ-Basis) **können ergänzt werden**. Zur Anreicherung dieser Informationen an die Standortinformationen werden verschiedene Verfahren eingesetzt. Das genaue Verfahren ist sehr komplex, basiert aber im Grundsatz auf **technisch-organisatorischer Trennung** und der **Verwendung verschiedener Verschlüsselungsmechanismen**.

Mehrstufiges Anonymisierungsverfahren: Schritt 1 – Segmentierung.

Im ersten Schritt werden die Datensätze zunächst in ihre kleinstmöglichen Bestandteile segmentiert und der direkte Personenbezug entfernt.

Anschließend werden diese Segmente auf verschiedene und unabhängige Bereiche in getrennte Hoch-sicherheitszentren unterschiedlicher Unternehmen verteilt („Zonenkonzept“).



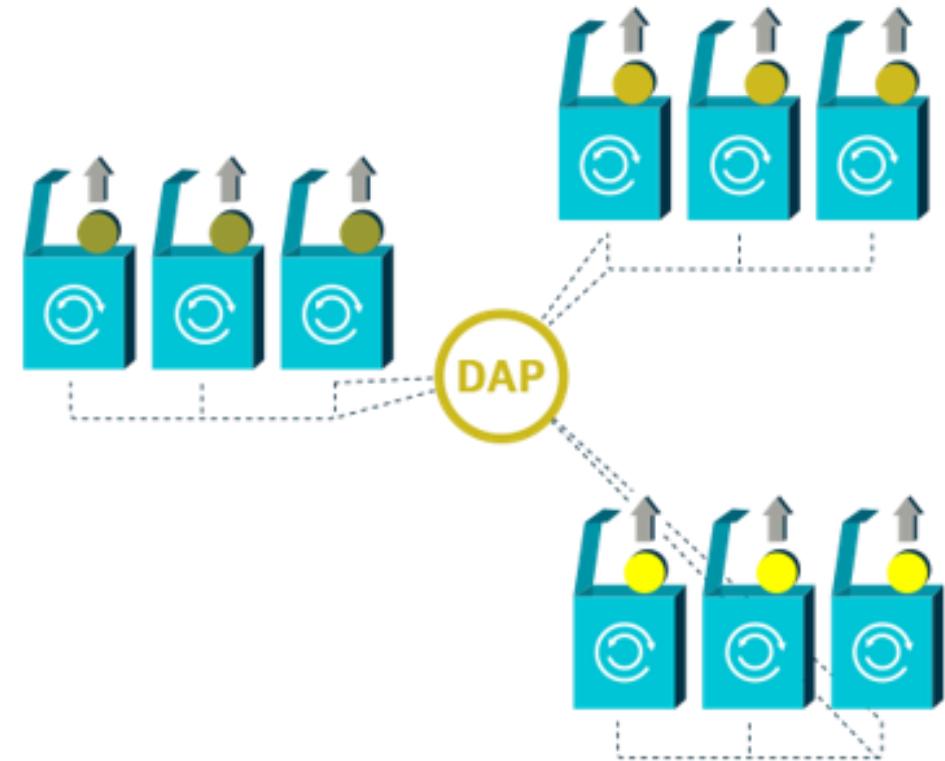
Mehrstufiges Anonymisierungsverfahren: Schritt 2 – Transformation.

In den verschiedenen Zonen werden die Segmente nach vordefinierten Schritten verarbeitet. Die vorhandenen Merkmale der Daten werden dabei per Zufallsprinzip entweder verändert oder sogar verworfen.

Hat eine Person z.B. einen sehr abgelegenen Wohnort oder Arbeitsplatz, an dem sonst fast keine Daten anderer Personen anfallen, könnte sie auch wenn ihre Daten noch so sicher verhasht sind - durch einfaches Beobachten der Situation vor Ort identifiziert werden. Um dies zu vermeiden, werden entsprechende Informationen unterdrückt oder verändert.

Auch k-Anonymität wird in diesem Schritt sichergestellt.

Die einzelnen Zonen, die die jeweiligen Informationen speichern, wissen zudem nicht, welche Informationen sie aufbewahren.



Mehrstufiges Anonymisierungsverfahren: Schritt 3 – Aggregation und Extrapolation.

Informationen werden nur dann ausgegeben, wenn diese gleichartige Merkmale einer größeren Gruppe beschreiben. Diese Gruppe bildet die Grundlage für eine Hochrechnung auf die gesamte Kundenbasis und – sofern sinnvoll für den konkreten Anwendungsfall – auf die Gesamtbevölkerung.

Auf diese Weise wird eine statistische Aussage generiert, wie z.B. „Am 06.12.2022 um 17 Uhr befanden sich am Marienplatz in München ziemlich genau XX (aber mindestens 5) Frauen im Alter von 21 bis 30 Jahren.“

Kurzzeitberechnungen erfolgen auf anonymen Echtdaten, Langzeitberechnungen auf Kurzzeitstatistiken (strikte Trennung zwischen beiden Stufen).



Mehrstufiges Anonymisierungsverfahren: Unverhältnismäßigkeit durch technisch-organisatorische Maßnahmen.

Technische Maßnahmen (Beispiele)

- Verwendung symmetrischer asymmetrischer Verschlüsselungsverfahren
- Vielfach segmentierte Verarbeitung und Speicherung der Daten
- Strikte technische Trennung von verarbeitenden Bereichen
- k-Anonymität und Verallgemeinerung von Eventdaten
- Veränderung von Zeitangaben bei beobachtbaren Ereignissen
- Real-Time-Verfahren (Flüchtigkeit)

Organisatorische Maßnahmen (Beispiele)

- Trennung in unterschiedliche Zonen (Zonenkonzept)
- Unterschiedliche juristische Personen
- Zusammenarbeit mit externem Analysepartnern bei der Erstellung der statistischen Analysen
- AV-Vertrag mit Dienstleistern, obwohl diese nur anonyme Daten verarbeiten; dabei u.a. Verbot der Verwendung der Daten zu anderen als den angewiesenen Zwecken mit Vertragsstrafe bewehrt

Unsere Kunden haben jederzeit die Möglichkeit, der Anonymisierung ihrer Daten in der DAP zu widersprechen.

Anwendungsfälle.

Die Mobilität der Bevölkerung sichtbar und nutzbar machen.

03

Insights durch anonymisierte Mobilfunkdaten.

- Mit der DAP kann o2 Telefónica das Mobilitätsverhalten unserer Kunden anonym darstellen und auf die Gesamtbevölkerung hochrechnen.
- Zahlreiche Anwendungsfälle, die sowohl der Gesellschaft als auch der Wirtschaft helfen, werden so ermöglicht.
- Gleichzeitig ist sichergestellt, dass die Rechte der Betroffenen hinreichend geschützt sind.
- Im Laufe der letzten Jahre hat o2 Telefónica zahlreiche unterschiedliche Anwendungsfälle umsetzen können – teilweise in Kooperation mit Städten, Forschungseinrichtungen oder ÖPNV-Betreibern.



Konkrete Anwendungsfälle (1/3).



Platzierung von Außenwerbung

Optimierung von Werbestandorten, um bestimmte Kundensegmente anzusprechen und digitale Anzeigen für verschiedene Nutzergruppen anzupassen.



Reduzierung von Feinpartikeln

Unterstützung von Städten mit datengestützten Analysen, um ihre Klima- und Luftreinhalteziele effizienter zu erreichen.



Städteplanung

Verständnis, welche Teile der Bevölkerung sich wie wohin bewegen, um z.B. den Bedarf an neuem öffentlichen Nahverkehr zu ermitteln.

Konkrete Anwendungsfälle (2/3).



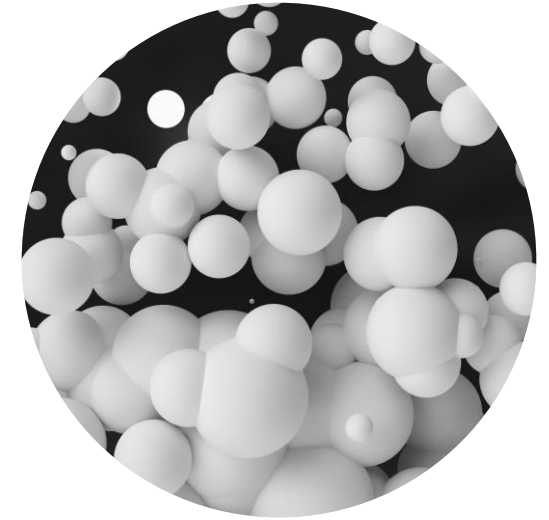
Verkehrsplanung

Verständnis der unterschiedlichen Mobilitätsformen verschiedener Bevölkerungsgruppen und Bewertung von Maßnahmen zur Kontrolle von Mobilitätsströmen.



Wetterbezogene Analysen

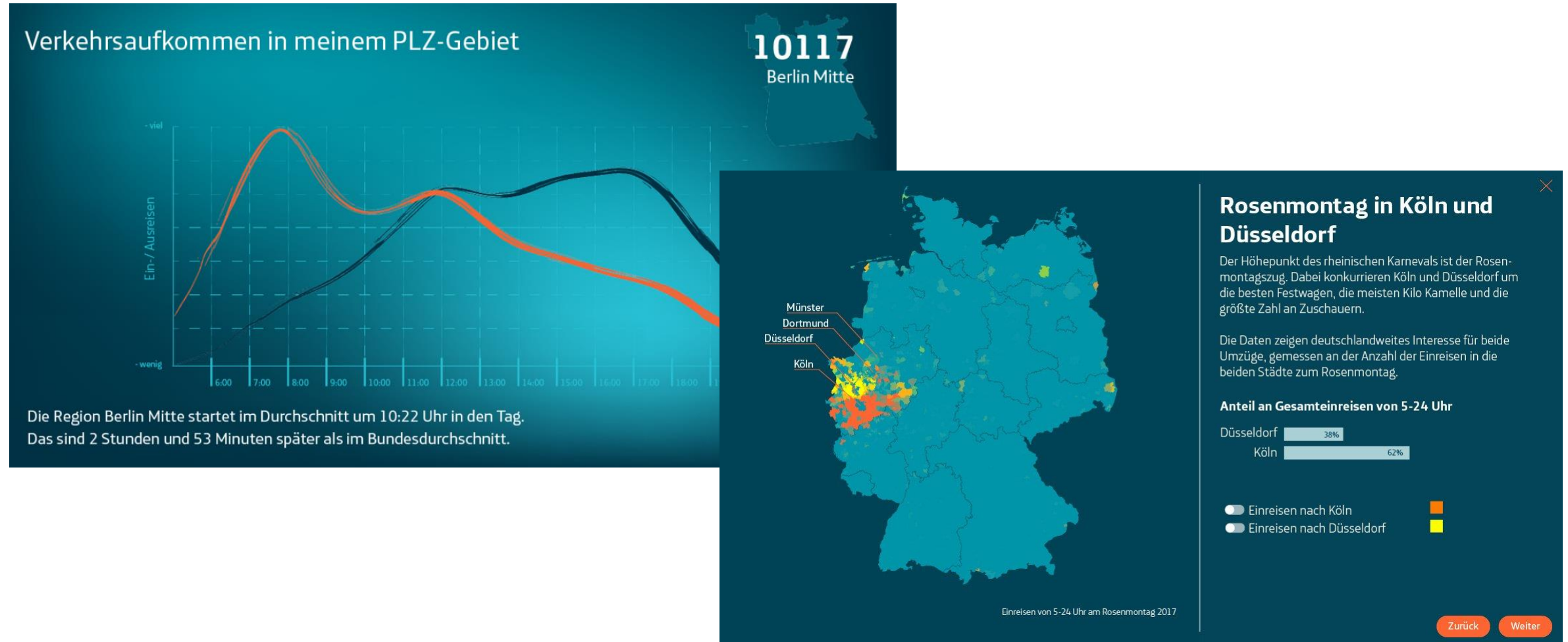
Verständnis der Auswirkung von Wetterphänomenen wie Stürme oder Starkregen auf das Reiseverhalten der Bevölkerung. (Selbiges gilt für Volksfeste, Fussballspiele, etc.)



Einsatz in der Pandemie

Überprüfung der Wirksamkeit des Lockdowns im Rahmen der COVID-19-Pandemie im Hinblick auf die Verringerung der Mobilität der Gesamtbevölkerung.

Konkrete Anwendungsfälle (3/3).



Kontakt.

Vielen Dank für Ihr Interesse!

Fragen gern an timo.wilken@telefonica.com oder
kontaktieren sie mich hier:





Telefónica