

---

# Datensystematik

Ein Glossar für die Nationale Dialogplattform Smart Cities

---



## Einleitung

Ziel der integrierten nachhaltigen Stadtentwicklung ist es, die enge Verknüpfung wirtschaftlicher, ökologischer, klimatischer, sozialer und demografischer Herausforderungen in ihrer Komplexität zu erkennen, integrierende Lösungsansätze zu entwickeln und umzusetzen. Entsprechend sollten zur Ausgestaltung einer zielgerichteten und funktionsfähigen Datengovernance in Kommunen, Daten aus allen Bereichen der Stadtentwicklung in den Blick genommen werden. Dazu gehören zum Beispiel:



Darüber hinaus liegen Daten innerhalb der Stadtentwicklung in vielerlei Arten und Weisen vor. Das Glossar bietet einen allgemeinen Überblick über die Typologie der Datenarten mit exemplarischen Anwendungsfällen aus der digitalen Stadtentwicklung.

## Typologie der Datenarten

| Merkmal                      | Ausprägung  | Definition   |
|------------------------------|---|--|
| <b>1. Datenurheberschaft</b> | <b>Amtliche Daten</b>   | Amtliche Daten umfassen alle Daten, die innerhalb des administrativen-politischen Systems zur Verfügung stehen.  |
|                              | <b>Forschungsdaten</b>  | In wissenschaftlichen Untersuchungen und Experimenten generierte Daten sind Forschungsdaten.   |
|                              | <b>Daten von Organisationen</b>   | Daten, die in Vereinen oder zivilgesellschaftlichen Organisationen anfallen.   |
|                              | <b>Unternehmensdaten</b>  | Unternehmensdaten sind alle in einem Unternehmen anfallenden Daten. Diese können eigens erhobene oder auch externe Daten sein.   |
|                              | <b>Private Daten</b>  | Persönliche Daten umfassen Daten, die beim Individuum selbst liegen.   |
| <b>2. Format</b>             | <b>Analoge Daten</b>  | Analoge Daten sind Daten, die auf einem analogen Weg Informationen vermitteln, wie z.B. eine analoge Uhr oder eine Tachoanzeige im Auto. Analoge Daten werden durch kontinuierliche Funktionen dargestellt, das heißt, dass es beliebig vielen Zwischenwerte gibt.   |
|                              | <b>Digitale Daten (digital data)</b>                                      | Digitale Daten sind alle Arten von Nachrichten und Informationen, die aus einer Folge von diskreten, also abgegrenzten nicht kontinuierlichen Werten bestehen und dadurch maschinell zu verarbeiten sind. Digitale Daten können in unterschiedlichen Formaten vorliegen, wie z.B. alphabetische Daten, numerische Daten sowie alphanumerische Daten. |
| <b>3. Veredelungsgrad</b>    | <b>Rohdaten/ Primärdaten (primary data)</b>                               | Rohdaten oder auch Primärdaten sind Daten, die unmittelbar erhoben wurden, jedoch noch keiner Analyse oder Operationalisierung durch eine Person, Software oder Maschine unterzogen wurden. Messdaten, die einem einzelnen Element einer Stichprobe zugeordnet sind, werden „ <b>Individualdaten</b> “ genannt.                                      |
|                              | <b>Verarbeitete Daten/ Sekundärdaten (processed data/ secondary data)</b> | Verarbeitete Daten/ Sekundärdaten umfassen Rohdaten bzw. Primärdaten, die verarbeitet wurden, um Informationen aus dem Material zu erhalten. Dabei wird zwischen (a) aufbereiteten   |

|                  |                              |  |
|------------------|------------------------------|--|
|                  |                              | <p>Daten und (b) Daten als Produkt einer Analyse unterschieden.</p> <p>(a) Verarbeitete oder aufbereitete Daten wurden beispielsweise aggregiert oder einzelne Merkmale wurden transformiert/rekodiert. Bei aggregierten Daten spricht man auch von „<b>Big Data</b>“. Neben dem Aggregieren gibt es viele verschiedene Verarbeitungsarten, wie z.B. konsolidieren, sortieren, zensieren, prüfen, gewichten oder löschen. Die Aufbereitung von Daten stellt jedoch noch keinen Analyseschritt dar.</p> <p>(b) Sekundärdaten können ein Produkt einer Datenanalyse, wie z.B. einer Clusteranalyse sein. Hierbei werden aus Rohdaten oder aufbereiteten Sekundärdaten mittels einer Clusteranalyse klar abgrenzbare Gruppen identifiziert.</p> |
|                  | <b>Metadaten (meta data)</b> | Metadaten sind strukturierte Daten, die Informationen über gesammelte Daten enthalten, die Auskunft über die Merkmale der Datensätze geben.  |
| <b>4. Zugang</b> | <b>Open data</b>             | Open data sind Daten, die ohne Einschränkungen zugänglich sind und allgemein weiter genutzt, verarbeitet oder verbreitet werden dürfen. Open data kann von verschiedenen Akteuren und in verschiedenen Formaten zur Verfügung gestellt werden. Häufig werden sie auf Plattformen bereitgestellt. Open data sind frei verfügbare Daten, jedoch können auch hier ggf. Einschränkungen für die weitere Verwendung und Nutzung vorliegen.  |
|                  | <b>Shared data</b>           | Shared data sind Daten, die gemeinsam genutzt werden können, wenn bestimmte Voraussetzungen erfüllt sind, z.B. kommerzielle Daten, die gegen einen monetären Wert (Lizenzkauf oder Paywall) zur Nutzung oder Weiterverwendung erworben werden können oder auch Daten des öffentlichen Sektors sind, die durch Anfragen im Sinne des Informationsfreiheitsgesetzes eingesehen werden können. Neben dem Zugang kann die Nutzung von shared data einer Einschränkung unterliegen.   |

|                                     |  |   |
|-------------------------------------|--|---|
|                                     | <b>Closed data</b>                           | Closed data sind Daten, die nicht frei zugänglich sind und nur durch ihren Besitzer/in eingesehen werden können, z.B. interne Daten, die vorliegen, aber aus verschiedenen Gründen nicht zugänglich sind. Ein Beispiel wären polizeiliche Ermittlungsdaten.   |
| <b>5. Dynamik der Daten</b>         | <b>Stammdaten „Sonderfall“ (master data)</b> | Stammdaten sind Daten, die eine hohe Statik aufweisen und sich selten verändern. Sie besitzen eine lange Gültigkeit sowie Beständigkeit, können sich jedoch u. U. auch verändern. Beispiele für Stammdaten wären Straßennamen oder Fingerabdrücke.  |
|                                     | <b>Statische Daten (static data)</b>         | Statische Daten stellen einen vergangenen Zustand dar. Sie umfassen Informationen für einen definierten Zeitraum.   |
|                                     | <b>Dynamische Daten (data stream)</b>        | Dynamische Daten sind Daten, die einen Live-Status wiedergeben und sich kontinuierlich aktualisieren (data stream). Sie ermöglichen ein unmittelbares Feedback, wie z.B. Echtzeitdaten oder Bewegungsdaten.   |
|                                     | <b>Manuell erhobene Daten</b>                | Manuell erhobene Daten sind Daten, die einzeln und nicht maschinell erhoben werden, wie z.B. Sozialdaten durch eine Befragung oder Dokumentendaten sowie Unternehmensdaten, die durch Behörden oder Unternehmen erhoben werden.   |
| <b>6. Art der Erhebung</b>          | <b>Maschinell erhobene Daten</b>             | Maschinell erhobene Daten können z.B. Sensor-Daten sein, die über Sensoren erhoben werden und Informationen über ihr Umfeld liefern. Wenn die Sensoren untereinander kommunizieren, handelt es sich um machine-to-machine-data.   |
|                                     | <b>Crowd-sourced data</b>                    | Crowd-sourced data entsteht mit Hilfe einer großen Gruppe von Akteuren. Mit diesem Ansatz lassen sich kostengünstig große Datensätze erzeugen. Crowd-sourced data kann manuell und maschinell erhoben werden. Ein Beispiel für die manuelle Erhebung wäre ein dezentrales Netzwerk von Wetterstationen auf privaten Grundstücken. |
| <b>7. Grad der Personalisierung</b> | <b>Personenbezogene Daten</b>                | Personenbezogene Daten umfassen alle Daten, die sich auf eine eindeutig identifizierbare Person   |

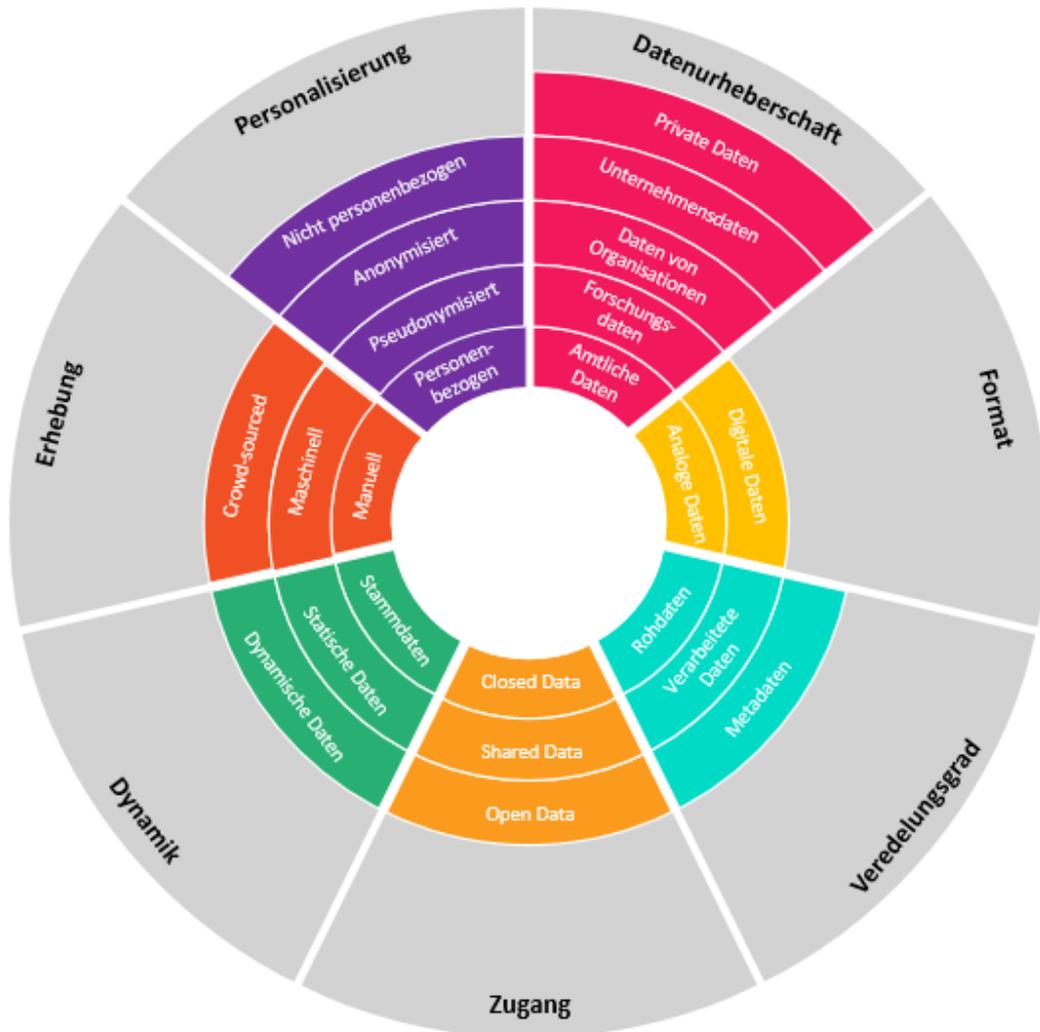
---

|   |   |
|---|---|
| <b>(personal data)</b>                                  | beziehen. Es muss eine klare Identifizierbarkeit und Beziehung zwischen der Information und der Person vorliegen.   |
| <b>Pseudonymisierte Daten</b>                           | Bei pseudonymisierten Daten werden Namen und andere personenbezogene Merkmale durch ein Pseudonym ersetzt. Dies kann ein Code oder eine zufällig gewählte Buchstaben- oder Zahlenkombination sein. Im Gegensatz zur Anonymisierung bleiben bei der Pseudonymisierung einige Bezüge zwischen den Informationen und den Datensätzen erhalten. |
| <b>Anonymisierte Daten</b>                              | Anonymisierte Daten umfassen Daten, die nicht ihrer Quelle zugeordnet werden können, sodass diese anonym bleibt. Sie stellen nicht-personenbezogene Daten dar.  |
| <b>Nicht-personenbezogene Daten (non-personal data)</b> | Nicht-personenbezogene Daten sind Daten, die sich nicht einer identifizierbaren Person zuordnen lassen und keine personenbezogenen Daten enthalten, wie z.B. Wetterdaten oder anonymisierte Daten.  |

---



Abbildung 1: Typologie Datenarten



## Beispiele für Datenarten in der digitalen Stadtentwicklung

Ein Anwendungsfall stellt das **Luftgütemessnetzwerk** dar. Dieses dient:

- der Berechnung von Kennwerten der Luftverschmutzung zur Beurteilung der Luftqualität anhand von Grenz- und Leitwerten,
- der Ermittlung der Schadstoffbelastung für verschiedene Genehmigungsverfahren,
- der Ursachenermittlung der Luftverunreinigung,
- dem Verfolgen der Wirksamkeit von Maßnahmen zur Luftreinhaltung und
- der Information der Öffentlichkeit.

Hierfür werden mit einer Granularität von fünf Minuten maschinell und digital Messwerte für relevante in der Luft befindliche Stoffe erhoben und verarbeitet. Zu den Messwerten zählen die Konzentration von Schwefeldioxid (SO<sub>2</sub>), Stickstoffdioxid (NO<sub>2</sub>), Kohlenmonoxid (CO), Benzol und Ozon (O<sub>3</sub>) sowie Feinstaub (PM<sub>10</sub>) und Ruß. Nach der Messung und Aufarbeitung der Daten (Konsolidierung, Prüfung auf Sinnhaftigkeit, etc.) werden die Daten auf einer Seite der OpenData-Initiative der Stadt als Echtzeit- sowie auch als historische Daten der Öffentlichkeit in leicht verständlicher Form zur Verfügung gestellt.

Abbildung 2: Datenarten in der Smart City – Anwendungsfall „Luftgütemessnetzwerk“



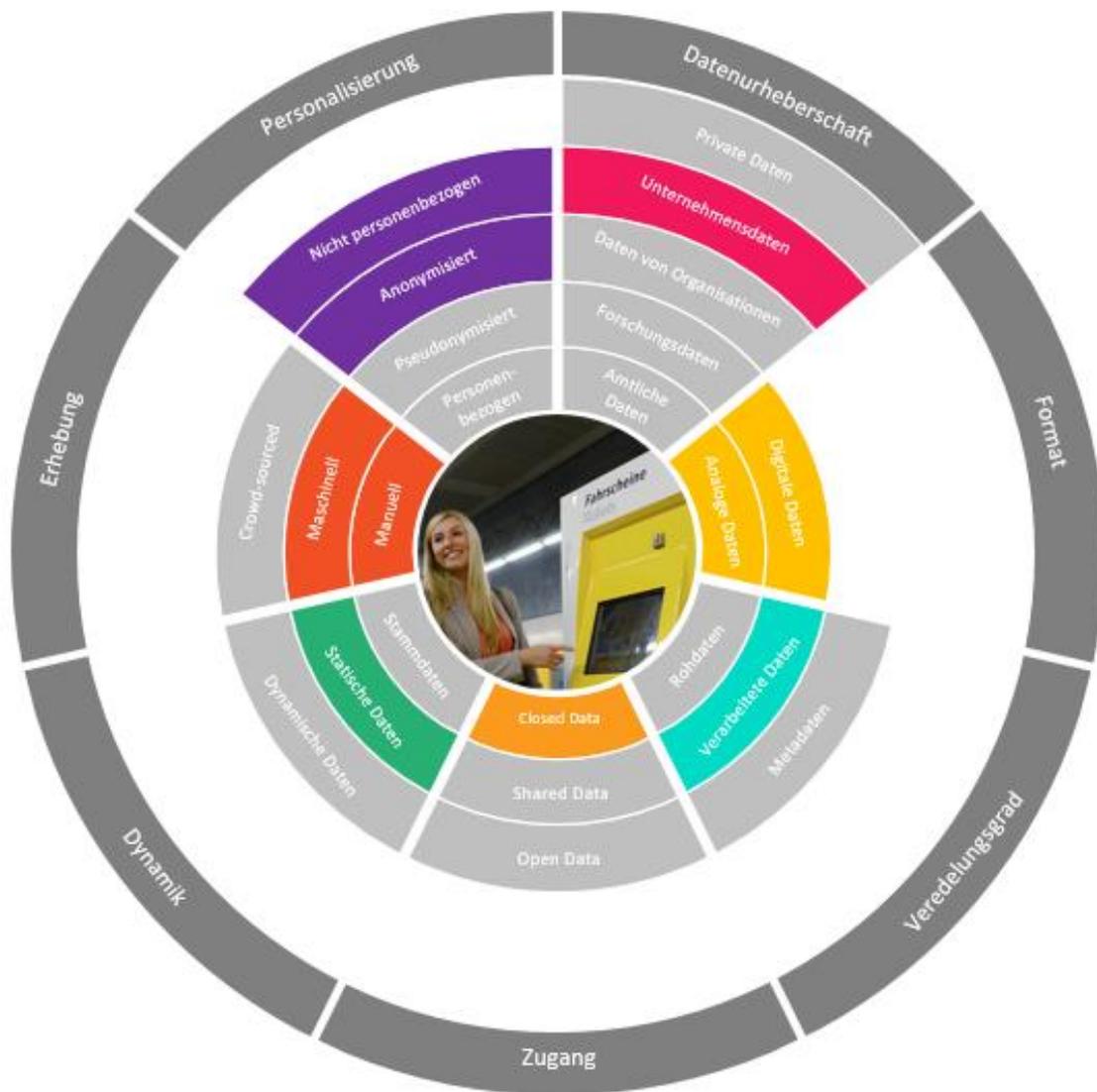
Ein weiterer Anwendungsfall ist die **Fahrgasterhebung öffentlicher Verkehrsbetriebe**. Dieser dient:

- der Bestimmung der Auslastung einzelner Linien zu vorher festgelegten Zeiten („Rush-Hour“, Wochenenden, Zeiten ohne starkes Berufspendleraufkommen)
- der Ermittlung von typischen Umsteigebewegungen,
- einem besseren Verständnis der Fahrkartenarten unter den verschiedenen Fahrgastgruppen,
- der qualitativen Erhebung von Kundenbedürfnissen z.B. nach Taktveränderungen, Haltestellenverlegung oder Linienausbau.

Hierfür werden innerhalb eines festgelegten Zeitraumes (z.B. einer Kalenderwoche) gezielt Passagiere in zu untersuchenden Regionen oder Netzabschnitten im Beförderungsmittel von beauftragten Teams angesprochen. Die Teilnahme ist freiwillig. Die Nutzer werden gebeten, Auskunft zu Einstiegs- und Ausstiegsstation zu machen, mitzuteilen ob Ihre Fahrt ein Umsteigen erfordert, welche Art von Fahrausweis sie besitzen (Einzelfahrausweis, Wochen-, Monats- oder Jahreskarte). Darüber hinaus wird nach Problemen bei der Beförderung gefragt, wie zum Beispiel einer denkbar günstigeren Linienführung oder Taktung.

Die Datenerhebung im Fallbeispiel erfolgt manuell und analog auf Papierfragebögen und wird für die weitere Verarbeitung ergänzt durch maschinell und digital erhobene Daten von Ticketverkaufsautomaten und anderen Quellen der Verkehrsbetriebe. Da es sich um eine einzelne Erhebungsaktion handelt, sind die Daten im Vergleich zu Stammdaten oder Echtzeitdaten als statisch zu betrachten. Die Erhebung entspringt betriebswirtschaftlichen Zwängen und wird für die Optimierung des Betriebs genutzt. Die gesammelten Daten verbleiben daher im Verkehrsunternehmen.

Abbildung 3: Datenarten in der Smart City – Anwendungsfall „Fahrgasterhebung öffentlicher Verkehrsbetriebe“



---

## Quellen

European Union (2020): Towards a European strategy on business-to-government data sharing for the public interest. Final report prepared by the High-Level Expert Group on Business-to-Government Data Sharing. URL: <https://ec.europa.eu/digital-single-market/en/news/experts-say-privately-held-data-available-european-union-should-be-used-better-and-more>

Europäische Union (2018): VERORDNUNG (EU) 2018/1807 DES EUROPÄISCHEN PARLAMENTS UND DES RATES vom 14. November 2018 über einen Rahmen für den freien Verkehr nicht-personenbezogener Daten in der Europäischen Union. URL: <https://eur-lex.europa.eu/legal-content/DE/TXT/PDF/?uri=CELEX:32018R1807&from=EN>

Fraunhofer FOKUS im Auftrag des Bundesministeriums für Bildung und Forschung (2018): Urbane Datenräume – Möglichkeiten von Datenaustausch und Zusammenarbeit im urbanen Raum. URL: [https://www.iais.fraunhofer.de/content/dam/iais/pr/pi/2018/PI\\_20180629/UDR\\_Studie\\_062018.pdf](https://www.iais.fraunhofer.de/content/dam/iais/pr/pi/2018/PI_20180629/UDR_Studie_062018.pdf)

Information Age (2015): The three types of data that will need to work together to run the Smart City. URL: <https://www.information-age.com/three-types-data-will-need-work-together-run-smart-city-123460068/>

Innoscale (2020): Was sind Stammdaten eigentlich? URL: <https://innoscale.de/blog-was-sind-stammdaten-eigentlich/>

Open Data Institute: The Data Spectrum. URL: <https://theodi.org/about-the-odi/the-data-spectrum/>

Nesta (2020): Data Sharing Toolkit. URL: [https://media.nesta.org.uk/documents/Data\\_Sharing\\_Toolkit\\_1.pdf](https://media.nesta.org.uk/documents/Data_Sharing_Toolkit_1.pdf)

Precht, Meier, Kleinlein (1994): EDV Grundwissen. Eine Einführung in Theorie und Praxis der modernen EDV. 2. Auflage. URL: [http://web4.wzw.tum.de/dvs/edvgw/e01\\_7.htm](http://web4.wzw.tum.de/dvs/edvgw/e01_7.htm)

Statista: Definition Individualdaten. URL: <https://de.statista.com/statistik/lexikon/definition/67/individualdaten/>

URBACT (2018): SmartImpact Action Planning Network URBACT III Management & Governance of Urban Data Thematic Report. URL: [https://urbact.eu/sites/default/files/media/tr\\_urbandata\\_v1.0\\_may2018\\_3.pdf](https://urbact.eu/sites/default/files/media/tr_urbandata_v1.0_may2018_3.pdf)