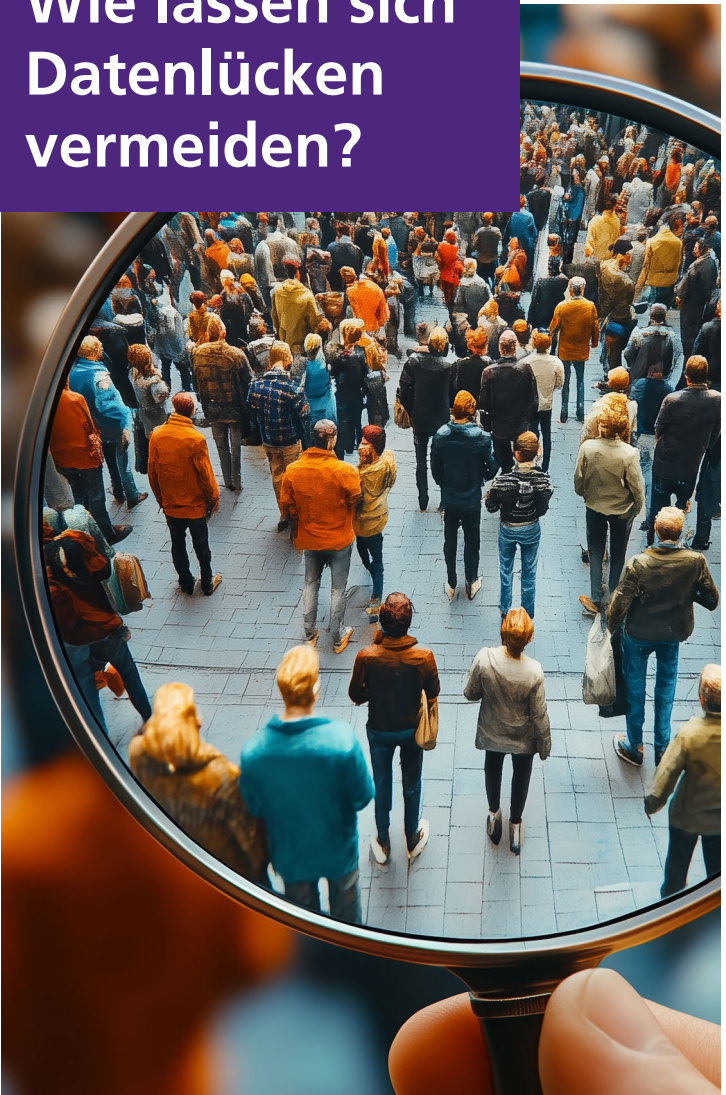


Leitfaden für Förderprojekte: Wie lassen sich Datenlücken vermeiden?



Was sind Datenlücken?

Datenlücken entstehen, wenn bestimmte Bevölkerungsgruppen oder Themen in Datensätzen fehlen oder nicht ausreichend repräsentiert sind. Das passiert zum Beispiel, weil nur bestimmte Menschen befragt werden oder die Fragen nicht für alle passen. Dadurch wird die Aussagekraft von Datenprojekten eingeschränkt und Innovationen werden ausgebremst. Themen wie soziale Vielfalt, Barrierefreiheit oder wirtschaftliche Unterschiede bleiben so oft unberücksichtigt.

Warum ist das Thema wichtig?

Eine große Herausforderung ist, dass manche Gruppen in Datensätzen unterrepräsentiert sind. Das betrifft nicht nur Merkmale wie Alter oder Herkunft, sondern auch verschiedene Lebenssituation, die bei standardisierten Erhebungen oft nicht erfasst werden. Das führt zu verzerrten Annahmen und verringert die Aussagekraft der Daten.

Warum fehlen manche Gruppen in den Daten?

Oft werden nur Menschen befragt, die leicht erreichbar sind. Manche Umfragen sind zudem zu kompliziert oder schwer verständlich. Es kann auch sein, dass manche Fragen nicht für alle relevant sind oder für einige Personen wichtige Fragen fehlen. Besonders die Bedürfnisse bestimmter Gruppen werden dadurch leicht übersehen.



Wer ist oft unterrepräsentiert?

- Ältere Menschen
- Menschen mit gesundheitlichen Einschränkungen
- Frauen
- Familien mit geringem Einkommen
- Menschen mit Migrationshintergrund
- Personen mit besonderen Mobilitätsbedürfnissen
(z. B. durch Kinderbetreuung oder Pflegeaufgaben)

Welche Folgen hat das?

Fehlende Daten führen dazu, dass Angebote und Infrastruktur nicht nach den tatsächlichen Bedürfnissen geplant werden können. Auch Fördergelder und Ressourcen könnten falsch eingesetzt werden. Dadurch können Maßnahmen, die auf den Projektergebnissen basieren, manche Gruppen unbeabsichtigt benachteiligen.

Was ist zu tun?

Dieser Leitfaden zeigt, welche typischen Probleme auftreten und wie Sie Daten inklusiv erheben und auswerten können. Dazu finden Sie Fragen zur Zusammensetzung der Stichprobe, zur Ansprache der Teilnehmenden, zur Auswertung und Interpretation sowie zur Dokumentation und Verbreitung der Ergebnisse.

Fazit

„Daten sollten die Realität möglichst genau abbilden.“

Wenn Gruppen fehlen oder nicht ausreichend vertreten sind, können falsche Entscheidungen getroffen werden. Um das zu vermeiden, sollten Befragungen verständlich sein und gezielt viele Menschen unterschiedlicher Gruppen einbezogen werden. So entstehen faire, realistische Lösungen. Nutzen Sie die bereitgestellten Ressourcen und teilen Sie Ihre Erkenntnisse, damit zukünftige Projekte noch inklusiver werden!

Kontakt: Dr. Katja Karrer-Gauß und Dr. Christina Wolking |
Institut für Innovation und Technik (iit) | wDRIVE@iit-berlin.de

Wie können Sie Datenlücken in Ihren Daten vermeiden?

Allgemein

Vielfalt bei der Zielgruppe sicherstellen

- ☐ Welche Personengruppen sind durch mein Projekt direkt oder indirekt betroffen?
- ☐ Sind alle relevanten Gruppen in meinen Daten vertreten?
- ☐ Eignet sich die Erhebungsmethode, um viele Menschen einzubeziehen?
- ☐ Können Nachbefragungen helfen, um fehlende Gruppen zu erfassen?
- ☐ Können Gewichtungen genutzt werden, um Repräsentativität sicherzustellen?



Interpretationen auf Verzerrungen prüfen

- ☐ Wird bei der Auswertung auf Unterschiede zwischen Gruppen geachtet?
- ☐ Gehen bei der Zusammenfassung (z.B. auf Haushaltsebene) wichtige Informationen verloren?
- ☐ Werden Ergebnisse mit bisherigen Daten verglichen und Abweichungen diskutiert?

Ergebnisse verständlich dokumentieren

- ☐ Sind offenen Datensätze so dokumentiert, dass sie leicht weiterverwendet werden können?
- ☐ Orientieren Sie sich an den Open-Data-Leitfäden der Bundesregierung?
- ☐ Wer kann Ihnen helfen, Ihre Datenerhebung zu verbessern?
- ☐ Nutzen Sie Feedback, um Ihre Ergebnisse allgemein verständlich zu formulieren?
- ☐ Machen Sie Einschränkungen bei Repräsentativität und Übertragbarkeit transparent?



Erkenntnisse weitergeben

- ☐ Werden die Ergebnisse veröffentlicht?
- ☐ Haben Sie Community-Plattformen oder Veranstaltungen genutzt, um Ihre Ergebnisse zu teilen?
- ☐ Wie bleiben Ihre Daten langfristig nützlich für andere Projekte?
- ☐ Sind die Ergebnisse praxisnah formuliert?

Befragungen

Unterschiedliche Personengruppen richtig ansprechen

- ☐ Ist die Einladung zur Befragung breit gestreut?
- ☐ Sind alle relevanten Gruppen über die vorgesehenen Kanäle erreichbar?
- ☐ Ist die Sprache einfach und verständlich?
- ☐ Ist die Befragung barrierefrei?



Relevante und passgenaue Fragen stellen

- ☐ Sind die Fragen für das Projekt relevant?
- ☐ Sind sie für alle Gruppen gleich relevant?
- ☐ Wurden Bedürfnisse einzelner Gruppen recherchiert und berücksichtigt?
- ☐ Gibt es Fragen, die bestimmte Gruppen ausschließen oder für einzelne Gruppen unterschiedlich wichtig sind?
- ☐ Fehlen Fragen, die für bestimmte Gruppen wichtig wären?
- ☐ Gibt es offene Fragen, um übersehene Perspektiven zu erfassen?



Daten differenziert auswerten

- ☐ Haben bestimmte Gruppen Fragen seltener oder gar nicht beantwortet?
- ☐ Gehen durch Zusammenfassungen wichtige Informationen verloren?
- ☐ Sind Analyseverfahren gewählt, die Unterschiede zwischen Gruppen zulassen (z. B. Clusteranalysen)?



KI-basierte Datenauswertungen

Eignung von Trainingsdaten kritisch prüfen

- ☐ Für welche Zwecke wurden Sekundärdaten ursprünglich erhoben? Sind sie für die eigenen Projektziele geeignet?
- ☐ Fehlen bestimmte Gruppen in den Datensätzen oder sind einzelne Gruppen unterrepräsentiert?
- ☐ Sind eigene, ergänzende Erhebungen oder weitere Datenquellen sinnvoll (z. B. Interviews, Fokusgruppen)?
- ☐ Sind synthetische Daten sorgfältig erstellt, dokumentiert und hinsichtlich ihrer Aussagekraft bewertet?

Datenlücken und Verzerrungen in KI-Modellen erkennen

- ☐ Wird bei bestehenden Modellen auf mögliche Benachteiligungen geprüft?
- ☐ Werden Methoden, Annahmen und Grenzen der KI-Modelle nachvollziehbar offengelegt?

Weiterführende Literatur

Kawgan-Kagan, I. (2023). The Gender Data Gap in Mobility - A systematic guide to understanding gender-sensitive data in mobility research. *Umweltpsychologie*, 1/2023 (Diversity of minds and people - Perspektivenvielfalt in der Umweltpsychologie). <https://doi.org/10.2440/015-0001>

Hessisches Ministerium für Wirtschaft, Energie, Verkehr, Wohnen und ländlichen Raum. (2024). Für wen planen wir? Datenlücken erfassen, Mobilität verbessern. Hessen Trade & Invest GmbH.

https://mobilitaetsplanung-hessen.de/wp-content/uploads/2025/01/Fuer_wen_planen_wir.pdf

Open Data – Handbuch für offene Verwaltungsdaten. https://www.bva.bund.de/SharedDocs/Downloads/DE/Behoerden/Beratung/Methoden/open_data_handbuch.pdf