RatSWD.

Rat für Sozial- und WirtschaftsDaten

Apps und Wearables: Chancen und Herausforderungen für die Wissenschaft

6 Output
6. Berufungsperiode

RatSWD.
Rat für Sozial- und WirtschaftsDaten

Datenerhebung mit neuer Informationstechnologie

Empfehlungen zu Datenqualität und -management, Forschungsethik und Datenschutz

GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

Mitwirkende bei der Erstellung

Mitglieder der AG Datenerhebung mit neuer Informationstechnologie

Prof. Dr. Thomas K. Bauer
Ruhr-Universität Bochum, RWI – Leibniz-Institut für Wirtschaftsforschung

Prof. Dr. Ulrich Ebner-Priemer
Karlsruher Institut für Technologie (KIT)

Prof. Dr. Michael Eid, *Vorsitz der AG*
Freie Universität Berlin

Prof. Dr. Anja S. Göritz
Albert-Ludwigs-Universität Freiburg

Dr. Cornelia Lange
Robert Koch-Institut (RKI)

Prof. Dr. Kai Maaz
Goethe-Universität Frankfurt, DIPF – Leibniz-Institut für Bildungsforschung und Bildungsinformation

Elke Nagel
Statistisches Bundesamt (Destatis)

Bertram Raum
Datenschutz-Experte

Dr. David Richter
Sozio-oekonomisches Panel (SOEP) am Deutschen Institut für Wirtschaftsforschung (DIW Berlin)

Prof. Dr. Mark Trappmann
Institut für Arbeitsmarkt- und Berufsforschung (IAB) der Bundesagentur für Arbeit (BA),
Otto-Friedrich-Universität Bamberg

Mitwirkende bei der Erstellung

Konsultation

Prof. Dr. Bernad Batinic
Johannes Kepler Universität Linz

Prof. Dr. Peter Dabrock
Deutscher Ethikrat

Prof. Dr. Eco de Geus
Vrije Universiteit Amsterdam

Prof. Dr. Björn Eskofier
Friedrich-Alexander-Universität Erlangen-Nürnberg

Prof. Matthias Mehl, Ph.D.
University of Arizona

Prof. Dr. Johannes Schöning
Universität Bremen

Prof. Kristof Van Laerhoven, Ph.D.
Universität Siegen

Prof. Dr. Cornelia Wrzus
Universität Heidelberg

Geschäftsstelle RatSWD

Dr. Mathias Bug

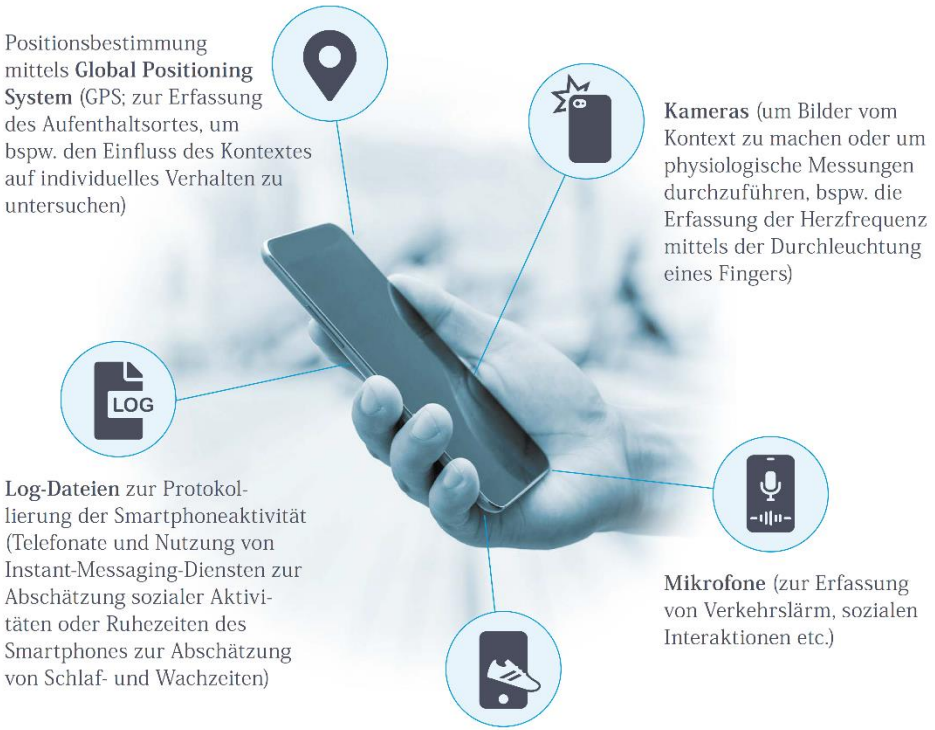
Dr. Tim Deeken

Dr. Nora Dörrenbächer

Bedeutung der Datenerhebung mit neuer Informationstechnologie

Smartphones – ein allgemein verfügbares Instrument der technisch unterstützten Erhebung von Forschungsdaten

Smartphones sind mit einer Vielzahl von Sensoren ausgestattet, die im Forschungsbereich der Sozial-, Verhaltens- und Wirtschaftswissenschaften Anwendung finden können. Dazu zählen unter anderem:



- Datenerhebung in „real life“
  - Ökologische Validität
- Datenerhebung in „real time“
  - Minimierung von Gedächtniseffekten
- Multimethodale Datenerhebung
  - Subjektive und objektive Maße
- Wiederholte Datenerfassung
  - Erfassung von Variabilität und Veränderung

Bedeutung der Datenerhebung mit neuer Informationstechnologie

- Zunahme von App-Entwicklungen und Start-Up-Unternehmen
- Umfangreiche Datenmengen
  - Kiukkonen et al. (2010),
    - 168 Personen / vier Monate
    - 15 Millionen Bluetooth-Scans
    - 13 Millionen WLAN-Scans
    - 15 Millionen GPS-Aufzeichnungen
    - 4 Millionen App-Nutzungs-Aufzeichnungen
    - 500.000 Akzelerometer-Messungen
    - 220.000 Sprachaufnahmen
    - 130.000 Sprachanrufe
    - 90.000 Textnachrichten
    - 28.000 Fotoaufnahmen
    - 2.000 Videoaufnahmen

Bedeutung der Datenerhebung mit neuer Informationstechnologie

- Visionen (Zitate aus Miller, 2012)
  - Psych apps for smartphones could become the default research method for most of psychology. (p. 233)
  - The question is not whether smartphones will revolutionize psychology but how, when, and where the revolution will happen. (p.234)
  - The scientific possibilities are limited only by our imaginations and by the technology - which will keep advancing faster than our imaginations can keep up. (p. 234)

Herausforderungen durch Datenerhebung mit neuer Informationstechnologie

- **Datenerfassung, Reliabilität und Validität**
- **Forschungsethik und Datenschutz**
- **Datenmanagement**

Herausforderungen durch Datenerhebung mit neuer Informationstechnologie

Datenerfassung, Reliabilität und Validität

- Smartphones sowie Wearables und andere Sensoren wurden originär nicht für die Forschung im engeren Sinne, sondern als Konsum-Produkte entwickelt und optimiert.
- Untersuchungen zur Messgüte und zur Validität dieser Technologien wurden bisher nur selten durchgeführt.
- Smartphones sowie die eingebauten Sensoren werden kontinuierlich weiterentwickelt und es werden Software-Updates durchgeführt, welche die Übertragbarkeit von Ergebnissen aus früheren Validierungsstudien in Frage stellen können.
- Inhalte solcher Veränderungen der Software bzw. die Algorithmen zur Gewinnung spezifischer Kennwerte werden von den Entwicklungs-Firmen häufig nicht offengelegt

Herausforderungen durch Datenerhebung mit neuer Informationstechnologie

Datenerfassung, Reliabilität und Validität

- Datenerhebung können nicht zu jeder Zeit und an jedem Ort erfolgen, da Signale nicht empfangen werden.
- Personen können nicht zu jeder Zeit Informationen zur Verfügung stellen.
  - z. B. Selbstberichtsdaten während der Steuerung eines Autos, beruflichen Sitzungen oder in Situationen, in denen die Privatheit gewahrt werden soll.
- Physiologische Messungen können mehrdeutig sein und erfordern zu ihrer korrekten Interpretation meistens die zusätzliche Erfassung des situativen Kontextes, in dem die Messung stattgefunden hat.
- Unterschiedliche Kompetenzen der Untersuchungsteilnehmenden beim Umgang mit neuer Informationstechnologie können darüber hinaus zu Selektionseffekten in Bezug auf die untersuchten Personengruppen führen.

Herausforderungen durch Datenerhebung mit neuer Informationstechnologie

Forschungsethik und Datenschutz

- Wie kann die Anonymität der untersuchten Personen gewahrt werden?
- Wie sollen Daten gespeichert werden, um die Gefahr einer Deanonymisierung auszuschließen?
- Wie können Personen selbstbestimmt darüber entscheiden, welche Daten sie zur Verfügung stellen und welche nicht?
- Welche Anforderungen ergeben sich aus der EU-Datenschutzgrundverordnung (DSGVO), insbesondere Art. 35 zur Datenschutz-Folgenabschätzung?
- Wie (ausführlich) soll eine informierte Einwilligung (informed consent) gestaltet werden?
- Welche Konsequenzen hat es, wenn strafrechtlich relevantes Verhalten aufgezeichnet wird?

Herausforderungen durch Datenerhebung mit neuer Informationstechnologie

Forschungsethik und Datenschutz

- Dürfen Daten anwesender Personen erhoben werden oder müssen diese konkret über die Ziele der Studie aufgeklärt und ihre informierte Einwilligung eingeholt werden?
- Wie können die Privatheit und Anonymität anwesender Personen gewahrt werden?

Herausforderungen durch Datenerhebung mit neuer Informationstechnologie

Datenmanagement

- Vielzahl an Datenpunkten
- Interoperabilität
- Open Science
  - Citizen Science: z. B. Open Human Projects
  - Datenbereitstellung für Sekundärnutzung
- Sicherung von Privatheit

Empfehlungen

- **Datenerfassung, Reliabilität und Validität**
  - 14 Empfehlungen
- **Forschungsethik und Datenschutz**
  - 35 Empfehlungen
- **Datenmanagement**
  - 5 Empfehlungen

Empfehlungen zur Datenerhebung

1) Datenerhebung erfordert eine umfassende **Dokumentation des Datenerhebungsprozesses**. Die Dokumentation bei Datenerhebung mit neuer Informationstechnologie sollte Informationen zu folgenden Aspekten umfassen:
   - Alle **verwendeten Sensoren und Softwares** (Hersteller, Typ, Herstellungsjahr, Softwareversion)
   - **Rohdaten**, die erhoben und gespeichert wurden
   - Einzelne **Schritte der Datenverarbeitung**, um zu abgeleiteten Merkmalen zu gelangen
   - **Kontextinformation** (z. B. Situation der Datenerhebung)

2) Wissenschaftlerinnen und Wissenschaftler sollten **abwägen, ob teurere wissenschaftstaugliche Produkte Konsum-Produkten vorzuziehen sind**. Die Erfassung von interessierenden Merkmalen ohne jeglichen Zugang zu den Rohdaten und ohne Kenntnisse über die im Produkt ablaufende Signalverarbeitung ist im Rahmen wissenschaftlicher Studien nur schwer zu rechtfertigen.

3) Sensordaten sollten möglichst als **Rohdaten in einem standardisierten Format** in ursprünglicher (nativer) Auflösung erfasst werden. Dabei empfiehlt es sich, verlustfreie Kompressionsarten zu benutzen.

4) Werden Sensordaten auf dem Gerät schon zu höherwertigen Informationen verarbeitet, sollte bekannt sein, auf welche Art und Weise dieses umgesetzt wird. Werden dazu **datenanalytische Verfahren**, wie z. B. maschinelle Lernverfahren, eingesetzt, sollten die Trainingsdaten und das jeweils verwendete Modell **bekannt** sein. Entsprechende Programmcodes und Programmversionen sollten zur Verfügung gestellt werden.

5) Hersteller sollten **Dokumentationen** aller von ihnen hergestellten Sensoren und Datenverarbeitungsalgorithmen **für alle Versionen der Sensoren archivieren**. Dies ist jedoch bei marktüblichen Konsum-Produkten nicht zu erwarten. Entsprechend sollten unbedingt bereits beim Kauf von Produkten die entsprechenden Datenblätter archiviert werden. Ebenso sollte versucht werden, Software- und Hardware-Updates zu protokollieren.

Empfehlungen zur Wahrung von Reliabilität

1) Informationen zur **Messgenauigkeit von Sensoren** sollten von Herstellern berichtet und in wissenschaftlichen Publikationen, in denen Sensordaten verwendet werden, **angegeben** werden.
2) In empirischen Anwendungen sollten **Maße der Messgenauigkeit** berichtet werden. Hierbei kann auf verschiedene Methoden (z. B. Paralleltestmethode, Retestmethode) zurückgegriffen werden.
3) Werden komplexe Daten (z. B. Audio-, Video-, Textdaten) z. B. durch Kodierverfahren reduziert, muss die **Güte der Reduktionsverfahren** (z. B. durch Interrater-Reliabilität bzw. Maße der Beurteilungsübereinstimmung) **dokumentiert** werden.

Empfehlungen zur Beurteilung der Validität

Konstruktvalidität

1) Für Sensoren, die in wissenschaftlichen Studien eingesetzt werden, sollte anhand von Validierungsstudien die **Konvergenz mit Gold-Standard-Methoden** nachgewiesen werden.

2) Durch Sensoren gewonnene Ergebnisse sollten in **theoriegeleiteten Validierungsstudien** (u. a. mit Rückgriff auf andere Datenquellen wie z. B. Verhaltensbeobachtungen und Befragungen) auf ihre Gültigkeit überprüft werden.

3) Bei mehrdeutigen Signalen sollten – soweit dies mit forschungsethischen und datenschutzrechtlichen Standards zu vereinbaren ist (siehe Kapitel 4) – durch die Hinzunahme weiterer Methoden sowie der Erfassung des Kontextes der Datenerhebung die **Gültigkeit der Schlüsse** nachgewiesen werden. Dies betrifft in besonderem Maße die Erhebung von physiologischen Prozessen im Alltag.

4) Werden Daten anhand von Algorithmen generiert oder ausgewertet, sollte in technischen Validierungsstudien die **Korrektheit der Algorithmen** aufgezeigt werden. Zudem sollten diese Daten mit realen Daten („ground truth“) abgeglichen werden. Sofern möglich, sollte dieser Abgleich im Rahmen jeder Datenerhebung (anhand von Teilstichproben) durchgeführt und dokumentiert werden. Ist dies in realen Anwendungen nicht möglich, sollte dies zumindest in Pilotstudien erfolgt sein.

Externe Validität

5) **Datenausfälle durch selektive Situationsstichproben** müssen dokumentiert und – sofern möglich und sinnvoll – anhand geeigneter Methoden rekonstruiert und korrigiert werden. Hierzu bieten sich verschiedene Vorgehensweisen an:

   a) Eine **nachträgliche Befragung der Probanden** (z. B. auch anhand entsprechend implementierter Fragen auf dem Smartphone) kann Aufschluss über die Gründe fehlender Daten geben und gegebenenfalls ermöglichen, fehlende Daten zu rekonstruieren. Eine derartige Befragung kann jedoch von den Befragten negativ bewertet werden.

   b) Eine Abschätzung der Selektivität der erfassten Situationen ist eventuell über einen **Vergleich** der Auftretenshäufigkeiten dieser Situationen in Datensätzen möglich, die **mit unterschiedlichen Methoden** erhoben wurden (Mobile Sensing, Experience Sampling, Feldbeobachtung, Fragebogen). Studien, die dies systematisch vornehmen, sind zwar extrem selten, jedoch finden sich Beispiele in der Forschung zu illegalen Rauschmitteln (Linas et al. 2016).

   c) Falls durchführbar (abhängig von der Belastung der Teilnehmenden und den vorhandenen Ressourcen) kann der **Erhebungszeitraum verlängert** oder die **Erhebungshäufigkeit erhöht** werden. So könnten z. B. bei kurz andauernden Situationen alle 60, 30 oder 15 Minuten Daten erhoben werden. Auch können Maße der Akzeptanz bzw. Belästigung erhoben werden, um deren Einfluss auf die Befolgung der Datenerhebungsanweisungen (engl. Compliance) abzuschätzen. Generell dürfte eine hohe Compliance durch solche Verfahren erreicht werden (Trull und Ebner-Priemer 2013). Systematische Studien hierzu sind jedoch noch sehr selten (Stone et al. 2003).

   d) Um das Ausmaß fehlender Daten abzuschätzen, kann erfasst werden, **wie häufig** Teilnehmende Möglichkeiten der **Editierung und Zensierung** ihrer Daten genutzt haben.

   e) Durch entsprechende **Incentivierung** (z. B. Teilnahmehonorar) kann versucht werden, die Compliance zu erhöhen (Göritz 2014).

6) **Datenausfälle** durch **selektive Personenstichproben müssen dokumentiert** und – sofern möglich und sinnvoll – anhand geeigneter Methoden **rekonstruiert und korrigiert werden**. Hierzu kann auf verschiedene Ansätze, u. a. aus der Surveyforschung (Kreuter et al. 2018; Schupp und Wolf 2015), zurückgegriffen werden:

   a) Um die Selektivität der Teilnehmenden analysieren und durch statistische Verfahren (z. B. Gewichtung) korrigieren zu können, kann es, wie in Kausalanalysen üblich, hilfreich sein, die **Teilnehmenden aus Datenquellen (zufällig) auszuwählen**, die relevante Informationen über alle Teilnehmenden und Nichtteilnehmenden enthalten. Dies können entweder administrative Daten wie Bevölkerungsregister sein oder aber bestehende repräsentative Surveys. Der Vorteil der Stichprobenziehung aus Surveys (vgl. Kreuter et al. 2018) ist, dass dort Fragen eingebracht werden können, die wichtige Zielvariablen der späteren Erhebung mit neuer Technologie vorwegnehmen, wie z. B. deren Nutzungsintensität oder Variablen, die mithilfe der neuen Technologie gemessen werden sollen. Es lässt sich dann modellieren, ob die Teilnehmenden der Studie bei den zentralen Maßen der Studie von Nichtteilnehmenden abweichen.

   b) Daten von Teilnehmenden (und ggf. der Kontrollgruppe) können **mit externen Datenquellen oder Statistiken verglichen** werden.

   c) Daten von Teilnehmenden und Nichtteilnehmenden können **mit kleinräumigen kommerziellen Daten** (z. B. infas 360) **verknüpft** werden (im Surveykontext vgl. Sinibaldi/Trappmann/Kreuter 2014).

   d) Nichtteilnehmende können zu **Gründen der Nicht-Teilnahme befragt** werden.

   e) Wird die Rekrutierung für die Studie persönlich durchgeführt, so besteht die Möglichkeit, dass Recruiter **Informationen zu Teilnehmenden und Nichtteilnehmenden beobachten bzw. schätzen** (vgl. West 2013).

   f) Es muss erfasst werden, ob und in welchem Umfang Geräte (z. B. Smartphones, Fitness-Tracker etc.) mit anderen Personen geteilt werden. Hierzu bietet sich an:

      - Die **Einwilligungserklärung** für wissenschaftliche Untersuchungen sollte – in Abhängigkeit vom Analyseziel – das **Teilen der zur Erhebung genutzten Wearables explizit ausschließen**.
      - Sofern sinnvoll, können Teilnehmende befragt werden, ob und wenn ja in welchem Ausmaß Geräte geteilt oder fremdverwendet wurden. **Von mehreren Personen verwendete Geräte können ausgeschlossen** oder – je nach Analyseziel – besonders berücksichtigt werden.
      - Mit Hilfe von Algorithmen des maschinellen Lernens kann es möglich sein, **unterschiedliches Nutzungsverhalten** zu identifizieren und die erhobenen Daten damit unterschiedlichen Nutzenden zuzuordnen (Ochoa/Bort/Porcar 2018). Dabei sind datenschutzrechtliche Vorgaben zu beachten (vgl. Kapitel 4).

Empfehlungen in Bezug auf Anonymität

1) **Anonymisierungs- und Pseudonymisierungsstrategien** sind grundsätzlich anzuwenden. Danach sind Daten nach dem Erreichen des Forschungszwecks grundsätzlich zu anonymisieren und bis dahin so zu speichern, dass Informationen durch Pseudonymisierung nicht zugeordnet werden können. (§ 27 Absatz 3 Bundesdatenschutzgesetz [BDSG]).
2) Solange eine Anonymisierung nicht sichergestellt bzw. erfolgt ist, müssen die rechtlichen Vorgaben der DSGVO umgesetzt werden. Insbesondere ist die **Notwendigkeit einer Datenschutz-Folgenabschätzung zu prüfen** und ggf. durchzuführen.
3) Daten sollten **nur auf sicherem Wege gespeichert, verarbeitet und transferiert** werden, z. B. unter Rückgriff auf Verschlüsselungstechniken (Brown/Blake/Sherman 2017; Carter et al. 2015).
4) Generell können One-Way Hashfunktionen zur Wahrung der **(formalen) Anonymität von Studienteilnehmenden** eingesetzt werden, so dass ein unmittelbarer Rückschluss auf eine bestimmte Person nicht länger möglich ist (siehe Naor und Yung 1989).
5) Von der Speicherung von Daten, welche die Reidentifizierung von Personen relativ leicht ermöglicht (z. B. Ortsinformationen) ist abzusehen, wenn sie nicht zur Beantwortung der konkreten Fragestellung benötigt werden. So sollten bspw. Geodaten direkt (z. B. per GPS, Wifi oder Mobilfunkdaten) oder indirekt (z. B. über Kontextinformation) nur dann erhoben werden, wenn sie zur Beantwortung der Fragestellung notwendig sind. Es sollten auch in anderen Datenquellen **„Spuren von Lokation" vermieden** werden (Geoinferencing bei Twitter-Daten), **sofern sie zur Beantwortung der Forschungsfrage nicht benötigt** werden (Johnson et al. 2016).
6) Es sollten **datenart-spezifische Möglichkeiten zur Anonymisierung** (die teils mit Informationsverlust einhergehen) **genutzt** werden, wie z. B. die Transformation (Verzerrung) von Stimmen bei Audioaufnahmen oder die Transformation von Koordinatensystemen (unter Beibehaltung der Topologie) bei räumlichen Informationen (Mainali/Shepherd/Petitcolas 2019).
7) **Daten** (wie z. B. Ortsdaten, Audiodaten) können **weniger präzise oder bereits weiterverarbeitet** (z. B. durch Behavioral Signal Processing) **abgespeichert** und die Rohdaten noch auf dem Gerät gelöscht werden, um die Möglichkeit der Reidentifikation zu verringern (z. B. Feng et al. 2018; Wyatt et al. 2011). Hier ergibt sich ein **Spannungsverhältnis mit den qualitätssichernden Empfehlungen** zur Konstruktvalidität (s. Empfehlungen in Kapitel 3.3.2) – es ist sinnvoll, einen Interessenabgleich durchzuführen und zu dokumentieren.
8) Daten mit **identifizierbaren Informationen** (z. B. volle Namen, Adressen, Kontonummern), die bspw. mittels Audio- oder Videoaufnahmen erhoben werden, sollten im Kodierprozess/weiteren Verlauf **gelöscht bzw. herauseditiert** werden (z. B. durch „beepen") (Robbins 2017).
9) **Erkennen datenverarbeitende Personen** (z. B. Kodierende) **die Studienteilnehmenden** (z. B. anhand ihrer Stimme), sollten sie die Weiterverarbeitung der Daten beenden und sie einer anderen Person übertragen (Robbins 2017).

Empfehlungen in Bezug auf den sozialen Kontext/beteiligte Dritte

1) **Personen**, die sich im regelmäßigen Kontakt mit den an einer Untersuchung teilnehmenden Personen befinden (z. B. Familie, befreundete Personen, Arbeitskolleginnen und -kollegen) und **von denen Daten erhoben werden, sind über die Datenerhebung** zu informieren und ihre Zustimmung ist einzuholen (Kelly et al. 2013).
2) Das **Aufzeichnungsrisiko sollte sichtbar gemacht werden**, um die **Privatheitserwartung bei Drittpersonen zu minimieren**. Hierzu können die an einer Untersuchung Teilnehmenden einen Recording-Anstecker tragen, der auf das Aufzeichnungsrisiko hinweist („Dieses Gespräch kann aufgezeichnet werden"), da dann von einer (passiven) Zustimmung der Probanden ausgegangen werden kann und die Aufzeichnung dann ethisch vertretbar ist (Mehl 2017; Robbins 2017). Alternativ sollten die an einer Untersuchung Teilnehmenden instruiert werden, das Aufzeichnungsrisiko Gesprächspartnern zu signalisieren (Robbins 2017).
3) **Geräte sollten ausgeschaltet werden, wenn** dies von Personen im sozialen Umfeld **gewünscht** wird (Brown/Blake/Sherman 2017; Robbins 2017).
4) Soweit **keine explizite Einwilligung** vorliegt, die personenbezogenen Daten nutzen zu dürfen, sollten **bei Audioaufnahmen die Stimmen dritter Personen** durch entsprechende Filter **verzerrt** werden.
5) Soweit **keine explizite Einwilligung** vorliegt, die personenbezogenen Daten nutzen zu dürfen, sollten **bei Videoaufnahmen die Gesichter von Personen verpixelt** werden.
6) Für die **Veröffentlichung von Daten**, die eine Identifikation Dritter erlauben, ist es notwendig, deren **Zustimmung einzuholen** (Kelly et al. 2013).
7) Personen des sozialen Umfelds, die anhand von Aufzeichnungen identifizierbar sind, muss die Möglichkeit eingeräumt werden, die **Zustimmung zur Datennutzung für bestimmte Zwecke jederzeit zu ändern und zu widerrufen** (Kelly et al. 2013).

Empfehlungen in Bezug auf Privatheit

1) Untersuchungsteilnehmende sollten präzise darüber informiert werden, in welcher Weise ihre Daten erhoben, gespeichert und weiterverarbeitet werden und welche Risiken in Bezug auf ihre Privatheit entstehen. Insbesondere sollten sie im Detail darüber aufgeklärt werden, **wie** sie **die Korrektur und Löschung von Daten fordern und Änderungen ihrer Zustimmungserklärungen vornehmen lassen** können (Kelly et al. 2013; Langheinrich und Schaub 2019) (s. Kapitel 4.2 – informierte Einwilligung).

2) Untersuchungsteilnehmenden sollte die **volle Kontrolle über ihre Daten** gegeben werden (informationelle Selbstbestimmung). So ist Personen die Möglichkeit einzuräumen, ihre eigenen Daten in zugänglichem Format zu erhalten (Art. 20 DSGVO), u. a. auch, um die Löschung spezifischer Sequenzen (**retroaktive Zensur**) fordern zu können (Kelly et al. 2013; Langheinrich und Schaub 2019; Mehl 2017).

3) Sofern möglich, sollten Untersuchungsteilnehmende **private Daten** (z. B. Audio- und Bildaufnahmen) **löschen können**, bevor die Forschungsdaten an die Forschenden weitervermittelt werden (Brown/Blake/Sherman 2017; Robbins 2017).

4) Das Aufnahmegerät sollte die **Möglichkeit einer proaktiven Zensur**, z. B. in Form eines Privacy-Buttons bieten, um die Datenaufnahme zu unterbrechen (Kelly et al. 2013; Langheinrich und Schaub 2019; Robbins 2017).

5) **Aufnahmegeräte sollten entfernt werden können**, wenn sich Untersuchungsteilnehmende dadurch in spezifischen Situationen wohler fühlen (Kelly et al. 2013).

6) Zum Schutz bei der Aufzeichnung von illegalem oder rufschädigendem Verhalten sollen **für die Teilnehmenden pro- und retroaktive Zensierungsmöglichkeiten gegeben sein**. Untersuchungsteilnehmende sollten darauf hingewiesen werden, dass (1) besonders schwerwiegende Straftaten den zuständigen Behörden gemeldet werden müssen und (2) bei allen Straftaten Studienunterlagen bei Ermittlungen der Strafverfolgungsbehörden diesen herausgegeben werden müssen.

7) **Kontinuierliche Aufzeichnungen** (z. B. bei Audioaufnahmen) **sollten vermieden werden**, sofern sie für die Beantwortung einer Forschungsfrage nicht notwendig sind. Kurze Segmente (30 Sekunden oder weniger) mit entsprechenden Zwischenpausen sind kontinuierlichen Aufzeichnungen vorzuziehen (sofern die Forschungsfrage dies zulässt) (Mehl 2017).

8) **Audio- und Bilddaten**, die es ermöglichen, Personen zu identifizieren, **sollten nicht weitergegeben werden, sofern keine explizite Zustimmung** der Personen, deren Daten erfasst wurden, vorliegen (z. B. in einer expliziten Zustimmungserklärung über die zukünftige Datennutzung) (Kelly et al. 2013).

9) **Rohdaten von Sensoren**, bei denen die Gefahr besteht, dass sie bei späteren Analysen neue personenbezogene Einblicke ermöglichen (z. B. Videodaten) sollten **nur in einer bearbeiteten Form** (z. B. mit einer geringeren Auflösung der Bilder oder mit weniger Frames pro Sekunde), **weitergegeben oder veröffentlicht** werden (z. B. Muaaz und Mayrhofer 2017; Ranjan und Whitehouse 2015).

10) **Sprachaufnahmen sollten dem Prinzip der Datenminimierung folgen**. Es sollen nur so viele Sprachaufnahmen vorgenommen werden, wie zur Verhaltensbestimmung benötigt werden (z. B. kurze Segmente [30 Sekunden oder weniger], keine hochfrequente Datenerhebung (Mehl 2017)).

11) Werden **Daten mit Smartphones erhoben**, so ist sicherzustellen, dass ein Zugriff auf private Daten des Smartphone-Besitzenden nicht möglich ist und nur die Daten weitervermittelt werden, deren Verarbeitung die Person zugestimmt hat.

Einwilligungsmodelle (Deutscher Ethikrat 2018: 183–185; DSGVO)

Blanko-Einwilligung

Untersuchungsteilnehmende stimmen einer inhaltlich unbestimmten zukünftigen Nutzung und der Weitergabe der Daten zu. Alle weiteren Entscheidungen darüber, was mit den Daten passiert, werden nicht mehr von den Untersuchungsteilnehmenden, sondern von der Verwaltung der Datenbanken getroffen.

Dynamische Einwilligung

Untersuchungsteilnehmende werden wiederholt kontaktiert, um ihre informierte Einwilligung zu einzelnen Fragestellungen einzuholen. Sie stehen in regelmäßigem Kontakt mit einer Datenbank (z. B. über Online-Plattform oder telefonische Kontakte).

Kaskaden- oder Meta-Einwilligung

Hierbei handelt es sich um eine Erweiterung der dynamischen Einwilligung. Die Einwilligung muss nicht zwangsläufig für jede Fragestellung neu eingeholt werden, sondern die Untersuchungsteilnehmenden können zwischen verschiedenen Optionen wählen. Neben dem klassischen dynamischen Einwilligungsmodell können sie bspw. ihre Zustimmung zu verschiedenen Kategorien von Forschungsfragen geben, ohne bei jeder einzelnen Forschungsfrage kontaktiert zu werden. Es ist auch möglich, die Entscheidung an andere Personen oder Einrichtungen (z. B. Expertengremien) treuhänderisch zu delegieren oder aber die eigenen Daten für jegliche Nutzung uneingeschränkt zur Verfügung zu stellen bzw. eine Teilnahme grundsätzlich abzulehnen. Die einzelnen Optionen können miteinander verknüpft werden bzw. zwischen den Optionen kann über die Zeit hinweg gewechselt werden.

Breite Einwilligung nach der DSGVO

Speziell für die wissenschaftliche Forschung wurde in Erwägungsgrund 33 der DSGVO die Möglichkeit einer “breiten” Einwilligung geschaffen (der sogenannte broad consent). Der europäische Verordnungsgeber geht dabei davon aus, dass es häufiger vorkommen kann, dass der Zweck der Verarbeitung personenbezogener Daten für Zwecke der wissenschaftlichen Forschung zum Zeitpunkt der Erhebung der personenbezogenen Daten nicht vollständig angegeben werden kann. Daher lässt er die Möglichkeit zu, dass Untersuchungsteilnehmende a) für bestimmte Bereiche wissenschaftlicher Forschung oder für Teile von Forschungsprojekten, b) in dem vom verfolgten Zweck zugelassenen Maß und c) unter Einhaltung anerkannter ethischer Standards der wissenschaftlichen Forschung eine “breite” Einwilligung erklären. Der “bestimmte Bereich” muss dabei einen Zusammenhang mit dem ursprünglichen Forschungsziel haben.

Empfehlungen in Bezug auf die informierte Einwilligung

1) **Teilnehmende** müssen über den Nutzen, die Risiken und die Form der Datenerhebung sowie über die Zwecke der Datennutzung, die Datenspeicherung und -weiterverwendung **in verständlicher Form aufgeklärt** werden, so dass sie informiert und aufgrund freier Entscheidung in die Verarbeitung ihrer Daten einwilligen können.
2) Die **Zustimmung** zur Datennutzung für bestimmte Zwecke **sollte jederzeit geändert und widerrufen werden können**.
3) Um sicherzustellen, dass **Einwilligungserklärungen** vollständig gelesen werden, sind **Kurzformen mit weiteren Erklärungen im Anhang denkbar**. In Internetstudien kann mit Videos gearbeitet werden, die Sinn, Vorgehen und Datenschutz der Studie erklären, wobei die Felder zur Bestätigung der Einwilligung erst erscheinen, wenn die Videos vollständig abgespielt wurden und geprüft wurde, ob der Ton angestellt war.
4) Um die Datennutzung langfristig sicherzustellen und das Selbstbestimmungsrecht der Teilnehmenden angemessen zu berücksichtigen, könnte ein **kaskadisches Einwilligungsmodell** einen Lösungsansatz darstellen.
5) Die **Freiwilligkeit der Teilnahme** muss stets gewährleistet bleiben. Der Abbruch der Studienteilnahme durch Probanden muss zu jeder Zeit möglich sein.
6) Den Probanden sollte die Möglichkeit zugesichert werden, **Einblick in die erhobenen Daten** zu nehmen und diese korrigieren zu dürfen.
7) Bei Video- und Tonaufnahmen ist die **Zustimmung Dritter** (z. B. Gesprächspartner) erforderlich und muss entsprechend eingeholt werden. Dies kann auch durch konkludente Handlungen erfolgen. Die **Einholung der Einwilligung muss allerdings dokumentiert werden**. Die Form der Dokumentation ist offen. Es können schriftliche Erklärungen, elektronische Erklärungen (E-Mail, Haken setzen in einem elektronischen Formular) oder auch Protokollierungen sein, die belegen, dass Dritte willentlich und informiert an der Studie teilgenommen haben.
8) **Apps sollten nur die für sie funktional notwendigen Berechtigungen** (Zugriff auf Adressbücher, Kamera, Mikrofon, Standortdaten etc.) **einfordern**. Es sollte keine Blanko-Freigabe erfolgen.

Empfehlungen zum Datenmanagement

1) Allgemeine **Standards für die Datensicherung**, die auch Interoperabilität sicherstellen, sollten weiterentwickelt werden.

2) Um die Qualität der Datenerfassung und -analyse bestimmen und Reanalysen angemessen durchführen zu können, sollten – wie bereits in Abschnitt 3.1 empfohlen – die Rohdaten sowie **alle Informationen über eingesetzte Geräte und weitere Datenverarbeitungsstrategien** (z. B. Algorithmen) archiviert und mit den Daten **zur Verfügung** gestellt werden.

3) Sollen Daten frei zugänglich, z. B. im Internet, zur Verfügung gestellt werden (z. B. in Open Repositories), muss sichergestellt werden, dass **eine Deanonymisierung nicht möglich** ist. Aufgrund der Vielzahl von Daten, die eine Deanonymiserung erleichtern, und der Unkenntnis darüber, inwieweit zukünftige datenanalytische Methoden weitere Möglichkeiten der Deanonymisierung eröffnen (siehe hierzu z. B. Rocher/Hendrickx/de Montjoye 2019), sollten solche Daten grundsätzlich nicht öffentlich zur Verfügung gestellt werden, es sei denn, Teilnehmende stimmen dem – nach Aufklärung über die Risiken – in einer informierten Einwilligungserklärung explizit zu.[6] Zu empfehlen ist die **Speicherung der Daten in Forschungsdatenzentren** mit limitiertem und geregeltem Datenzugriff vor Ort oder via Internet mit entsprechendem Passwort-Schutz. Die Passwörter könnten auch in den Forschungsdatenzentren verwaltet und nach Rücksprache mit den Autorinnen bzw. Autoren der Publikationen sowie mit vertraglicher Absicherung (Datennutzungsverträge, die Missbrauch ausschließen sollen) an die interessierten Forschenden herausgegeben werden.

4) **Datensicherung und -weiterverwertung** sollten in einer informierten **Einwilligungserklärung** geregelt werden. Hierzu kann auf die in Abschnitt 4 vorgestellten Verfahren, u. a. auch die **treuhänderische Verwaltung von Daten**, zurückgegriffen werden.

5) Strategien der Datenanalyse, die es ermöglichen, dass **Personen im Besitz ihrer Daten bleiben**, sollten weiterentwickelt werden. Rechenoperationen würden dann auf Daten zurückgreifen, die an verschiedenen Orten (bei den teilnehmenden Personen) gespeichert sind.