



4.1 Datenpflege

Daten managen – 4 Datenpflege und Schutz

Unsere Versorgungsforscherin wertet Patient:innendaten aus einem großen Universitätsklinikum aus, um verschiedene Behandlungswege zu vergleichen. Dabei stößt sie eines Tages bei ihrer Analyse auf widersprüchliche Ergebnisse. Statt voreilige Schlüsse zu ziehen, entscheidet sie sich, die Daten zu prüfen. Sie stellt fest, dass einige Datensätze unlogische und fehlerhafte Angaben zu Diagnosen und Behandlungszeiten enthalten. Unsere Versorgungsforscherin macht sich daran die Daten sorgfältig zu bereinigen: Sie überprüft Dubletten, klärt Kodierungsfehler und ergänzt fehlende Werte durch Rücksprache mit der Klinik. Nach der Bereinigung verändert sich das Bild deutlich: Plötzlich zeigen die Daten ein klareres Ergebnis, das es ermöglicht, gezielte Empfehlungen für eine bessere Patientenversorgung abzuleiten. Unsere Versorgungsforscherin erkennt, dass gute Forschung nicht nur viele Daten braucht, sondern vor allem gepflegte, konsistente (und gut dokumentierte) Daten.

Datenpflege

Datenpflege ist der Prozess der kontinuierlichen Überprüfung, Aktualisierung, Korrektur und Organisation von Daten in einem Informationssystem um deren Qualität, Verlässlichkeit und Nutzbarkeit sicherzustellen und zu bewahren. Das heißt die Daten werden kontinuierlich ergänzt und bereinigt, sowie Redundanzen und Fehler entfernt.

Datenqualitätsdimensionen

Damit Daten geteilt und wissenschaftlich (weiter) genutzt werden können, müssen sie dem erforderlichen Qualitätsstandard entsprechen. Dies betrifft nicht nur die Daten selbst, sondern auch die Metadaten (siehe für mehr Infos **3 Metadaten und Metadatenstandards**) und die Infrastruktur, über die Daten zur Verfügung gestellt werden.

Wang und Strong (1996) formulierten vier Dimensionen der Datenqualität:

- 1) Intrinsische Datenqualität:
 - a) Sind die Daten glaubwürdig und zuverlässig?
 - b) Sind die Daten fehlerfrei?
 - c) Wurden die Daten objektiv erhoben?
 - d) Hat die Datenquelle ein hohes Ansehen?
- 2) Kontextuelle Datenqualität
 - a) Bieten die Informationen einen Mehrwert?
 - b) Sind die Daten relevant?
 - c) Sind die Daten aktuell?
 - d) Liegen die Daten in einem angemessenen Umfang vor?
- 3) Repräsentationelle Datenqualität
 - a) Sind die Daten eindeutig auslegbar und nicht doppeldeutig?
 - b) Sind die Daten verständlich?
 - c) Sind die Daten übersichtlich und in einer gut erfassbaren Art dargestellt?
- 4) Zugriffsqualität
 - a) Sind die Daten bearbeitbar für die jeweiligen Nutzungszwecke?
 - b) Sind die Daten einfach zugänglich?

Prüfroutinen für Daten

Im besten Fall sind Daten natürlich schon von Beginn an fehlerfrei und benötigen keine Nacharbeit, erfüllen also direkt die

Gefördert durch:



Datenqualitätsdimensionen. Dadurch können Zeit und Kosten gespart werden. Das Konzept, solche Daten zu erhalten nennt sich auch „First-Time-Right-Prinzip“. Durch folgende Maßnahmen können von Beginn an Fehler in den Daten vermieden werden:

- Einheitliches System der Datenerstellung durch Standardisierungen (z.B. Datumsangaben im JJJJMMTT-System)
- Plausibilitätsprüfung (z.B. Altersangaben nur als Zahl und <150)
- Dublettenprüfung
- Mehraugenprinzip bei der Digitalisierung von analogen Daten (z.B. Eingabe von Papierfragebögen von zwei Personen unabhängig voneinander)

Natürlich ist es auch sinnvoll bestehende Daten, oder solche die nicht von einem selbst gespeichert wurden, auf Fehler zu prüfen.

Softwareempfehlung: Die open source Software [OpenRefine](#) kann Inkonsistenzen in großen Mengen finden und korrigieren, aber auch eine Prüfung durch Kolleg:innen kann Fehlermengen reduzieren.

Aktualisierung

Je nach Datensatz kann es üblich sein, dass immer wieder Änderungen an ihnen vorgenommen werden können: Möglicherweise kommen mehr Informationen dazu oder einzelne Angaben ändern sich. Bei solchen langlebigen Datensätzen ist es sinnvoll die Inhalte regelmäßig zu prüfen und ggf. anzupassen. Um hier den Überblick zu behalten, sollte immer unzweifelhaft sichtbar sein, welche Datenversion die aktuellste ist. Hierbei kann eine Versionskontrolle (siehe für mehr Infos **01 Datenstrukturierung**) helfen.

Löschen

Zur Datenpflege gehört auch unnötige Daten zu löschen oder Daten zu löschen, wenn die Aufbewahrungsfrist abgelaufen ist. Insbesondere doppelte Daten können zu Verwirrung und Fehlern führen und die kluge Entscheidung, welche Datenstände behalten werden müssen, ist essenziell. Einfaches Verschieben in den Papierkorb reicht hier nicht aus, die Daten müssen dauerhaft vernichtet werden. Dies ist z.B. durch spezifische Software möglich wie [Eraser](#), [WipeFile](#) oder [Permanent Eraser](#). Hardware, die nicht weiter benötigt wird, kann auch physisch vernichtet werden. Detaillierte Informationen über die richtige Vernichtung verschiedener Datenträger gibt die [DIN 66399](#).

Datenpflege in Arztpraxen

Spezifisch auf Gesundheitsdaten bezogen haben Patient:innen das Recht von ihren behandelnden Ärzt:innen die Berichtigung unrichtiger personenbezogener Daten zu verlangen, die nach objektiven Maßstäben nicht der Realität entsprechen (Art. 16 S. 1 DSGVO). Die nachträgliche Berichtigung oder Vervollständigung ist unter Beibehaltung der alten Angabe zu vermerken.

Patient:innen haben das Recht auf eine unverzügliche Löschung, wenn die Daten nicht mehr benötigt werden, die Einwilligung in die Verarbeitung widerrufen wurde, ein Widerspruch gegen die Verarbeitung erklärt wurde oder die Speicherung unzulässig ist. Ausnahmen bilden hier eine entgegenstehende vertragliche oder satzungsgemäße Aufbewahrungspflicht (Art. 17 Abs. 3 Buchst. b DSGVO) oder ein öffentliches Interesse im Bereich der öffentlichen Gesundheit oder der öffentlichen Forschung (Art. 17 Abs. 3 Buchst. c DSGVO i. V. m. Art. 9 Abs. 2 lit. i DSGVO und § 27 Abs. 2 BDSG.).

Quellenverzeichnis

Bundesärztekammer & Kassenärztliche Bundesvereinigung (2021). Hinweise und Empfehlungen zur ärztlichen Schweigepflicht, Datenschutz und Datenverarbeitung in der Arztpraxis. Deutsches Ärzteblatt. <https://doi.org/10.3238/arztebl.2021.ds02>. Abgerufen am 30.09.2025, von https://www.kbv.de/documents/infothek/rechtsquellen/weitere-vertraege/praxen/datenschutz-schweigepflicht/Empfehlungen_aerztliche_Schweigepflicht_Datenschutz.pdf

Krähwinkel, E., Langner, P., Lipp, R., & Pietsch, A. M. (2022, März 22). HeFDI Data OER: Forschungsdatenmanagement – eine Online-Einführung. Zenodo. <https://doi.org/10.5281/zenodo.6373596>

Wang, R. Y., & Strong, D. M. (1996). Beyond accuracy: What data quality means to data consumers. *Journal of management information systems*, 12(4), 5-33.

Koch, H., Schütze, B., Spyra, G., & Wefer, M. (2017). Datenschutzrechtliche Anforderungen an die medizinische Forschung unter Berücksichtigung der EU Datenschutz-Grundverordnung (DS-GVO). Abgerufen am 30.09.2025, von https://gesundheitsdatenschutz.org/download/forschung_ds-gvo.pdf

Weiterführende Info

Bundesärztekammer & Kassenärztliche Bundesvereinigung (2021). Hinweise und Empfehlungen zur ärztlichen Schweigepflicht, Datenschutz und Datenverarbeitung in der Arztpraxis. Deutsches Ärzteblatt. <https://doi.org/10.3238/arztebl.2021.ds02>. Abgerufen am 30.09.2025, von https://www.kbv.de/documents/infothek/rechtsquellen/weitere-vertraege/praxen/datenschutz-schweigepflicht/Empfehlungen_aerztliche_Schweigepflicht_Datenschutz.pdf

Reisswolf International GmbH (2025). Die Schutzklassen und Sicherheitsstufen der DIN 66399. Schritt für Schritt zur richtigen Datenträgervernichtung. Abgerufen am 30.09.2025, von <https://www.reisswolf.com/leistungsbereiche/aktenvernichtung-datenvernichtung/din-66399-schutzklassen/>

Koch, H., Schütze, B., Spyra, G., & Wefer, M. (2017). Datenschutzrechtliche Anforderungen an die medizinische Forschung unter Berücksichtigung der EU Datenschutz-Grundverordnung (DS-GVO). Abgerufen am 30.09.2025, von https://gesundheitsdatenschutz.org/download/forschung_ds-gvo.pdf

Das Material ist im Rahmen des Projekts „DIM.RUHR: Datenkompetenzzentrum für die interprofessionelle Nutzung von Gesundheitsdaten in der Metropole Ruhr“ (Förderkennzeichen: 16DKZ2008[A-F]) entstanden. Dementsprechend spiegelt das Material in Version 1.0 den Stand von 2025 wider.

„Daten managen - 4.1 Datenpflege und -schutz – Datenpflege“ von Anne Mainz und Sven Meister.

Dieses Werk und dessen Inhalt sind – sofern nicht anders angegeben – lizenziert unter CC BY 4.0. Ausgenommen aus der Lizenz sind die verwendeten Logos.

