

Was sind Daten?

Adrian Regez, Dr. Uwe Dirksen, Arbeitsgruppe KI der PHBern
11. Mai 2023

In einer datengetriebenen Welt, dem «Datapozän» wie der Journalist Adrian Lobe das heutige Zeitalter nennt [6], ist es wichtig, zu wissen, was Daten überhaupt sind. Der vorliegende Text versucht dies zu klären und stellt darüber hinaus ein Zyklusmodell vor, welches Daten in einem grösseren Kontext einbettet. Auch Künstliche Intelligenz (KI) beziehungsweise KI-Anwendungen verwenden Daten, was KI genau ist wird aber nicht hier, sondern in diesem Artikel [Link zu Artikel 2] beschrieben.

Der Begriff 'Datum'¹ (Einzahl von Daten) [8] kann je nach Blickwinkel verschiedene Bedeutungen haben. Zum Beispiel kann er in der Philosophie, Mathematik oder auch in der Informationswissenschaft unterschiedlich interpretiert werden. Im vorliegenden Zusammenhang liegt der Fokus hauptsächlich auf «kommunikativen Artefakten» [7], wie beispielsweise Texten, Bildern und Videos, die aus einer informationswissenschaftlichen Perspektive als Daten betrachtet werden.

Daten sind «kontextfreie Angaben», welche aus erfassten Zeichen beziehungsweise Signalen bestehen [8]. Diese **Zeichen** können beispielsweise (in einem Buch, Zeitschrift o.ä.) als Buchstaben, Zahlen oder Symbole auftreten, während im vorliegenden Kontext, **Signale** menschliche Sinneseindrücke «von Licht, Schall, Geruch, Geschmack und Berührung» beschreiben [5]. Auch Computer können solche 'Eindrücke wahrnehmen'², sie brauchen dazu Sensoren [5]. Sobald wir Menschen oder Computer beginnen, die 'Eindrücke' zu erfassen, entstehen Daten. Demnach bilden Zeichen & Signale das Rohmaterial ebensolcher Daten.

Wenn Daten weiterverarbeitet und analysiert werden, entstehen daraus **Informationen** [4, 8], die interpretiert werden und zusammen mit anderen Informationen **Wissen** schaffen [4]. Abbildung 1 zeigt das Zusammenspiel von Zeichen & Signalen, Daten, Informationen und Wissen (ZSDIW).

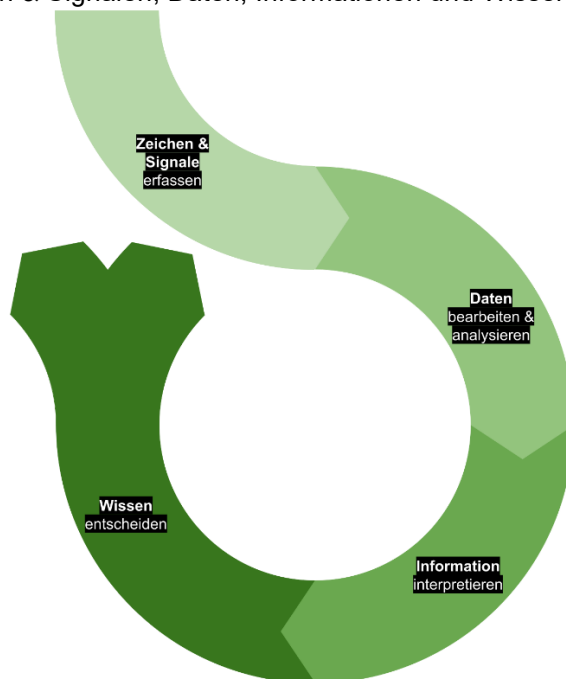


Abbildung 1: Zeichen-Signale-Daten-Informationen-Wissens-Zyklus (ZSDIW-Zyklus) in Anlehnung an verschiedene Interrelationskonzepte [1–4].

¹ Lateinisch = gegeben

² Der Computer kann nicht auf die gleiche Art und Weise 'wahrnehmen' wie wir Menschen: Zumal Computer mit Elektrizität arbeiten, benötigen sie Sensoren, die ihnen dabei helfen, beispielsweise Schall in elektrische Signale umzuwandeln, damit sie ebendiesen Schall auch erfassen können. Elektrische Signale können dabei zwei Zustände annehmen (binär) – Ein oder Aus. Dies ist das grundlegende Prinzip, auf dem Computer basieren. Diese Zustände werden oft als Nullen und Einsen dargestellt.

Nebst dem Zusammenspiel von ZSDIW zeigt das zyklische Modell von Abbildung 1 auch, dass Zeichen & Signale zwar die 'ursprüngliche' Grundlage von Daten sind, aber auch bereits generiertes Wissen als Datengrundlage verwendet werden kann. Ein Beispiel: Ein Unternehmen sammelt Kund*innen-daten, um das Kaufverhalten zu analysieren. Daraus kann beispielsweise Wissen über das (1) Alter der einzelnen Kund*innen und eine (2) Rangliste der meistverkauften Produkte generiert werden. Sowohl das Alter der Kund*innen ('Wissen 1'), als auch die Rangliste der meistverkauften Produkte ('Wissen 2') können wiederum als Daten verwendet und so analysiert sowie interpretiert werden. Als Resultat davon könnte beispielsweise das Wissen darüber entstehen, welche Kund*innen, in welchem Alter, welches Produkt kaufen.

Beim Wissen gilt es zwischen zwei Optionen zu entscheiden: Der Pfeil nach rechts in der Abbildung 1 bedeutet, dass das Wissen, wie eben beschrieben, wiederum verwendet wird, um schlussendlich neues Wissen zu generieren. Der Pfeil nach links stellt symbolisch einen Wissensoutput dar. Am vorigen Beispiel anknüpfend, könnte ein Output beispielsweise sein, dass das Wissen darüber welche Kund*innen, in welchem Alter, was kaufen, dem Verwaltungsrat des Unternehmens kommuniziert wird und daraus im Endeffekt eine neue Produktstrategie entworfen werden könnte.

Literaturverzeichnis

- [1] Russel L. Ackoff. 1989. From data to wisdom. *Journal of Applied Systems Analysis* 15, 3–9.
- [2] Dursun Delen. 2020. *Prescriptive analytics. The final frontier for evidence-based management and optimal decision making*. Pearson, New York.
- [3] Andreas Grillenberger. 2019. *Von Datenmanagement zu Data Literacy. Informatikdidaktische Aufarbeitung des Gegenstandsbereichs Daten für den allgemeinbildenden Schulunterricht*. Dissertation. Freie Universität Berlin.
- [4] Helmut Krcmar. 2015. *Informationsmanagement* (6. Aufl. 2015). Springer Berlin Heidelberg, Berlin, Heidelberg.
- [5] Anthony Liew. 2007. Understanding Data, Information, Knowledge And Their Inter-Relationships. *Journal of Knowledge Management Practice* 2, 7.
- [6] Adrian Lobe. 2019. *Speichern und Strafen. Die Gesellschaft im Datengefängnis*. C.H. Beck, München.
- [7] Jakob Voß. 2013. *Was sind eigentlich Daten?* DOI: <https://doi.org/10.18452/9038>.
- [8] Bernhard C. Witt. 2010. *Datenschutz kompakt und verständlich. Eine praxisorientierte Einführung* (2nd). Vieweg+Teubner, Wiesbaden.