

Die im Dunklen sieht man doch

Dunkle Daten und Datentracking in der Wissenschaft

In der Wissenschaft wurde schon sehr früh vom ›Data Deluge‹, also der gehörigen Flut an Forschungsdaten gesprochen.¹ Sie werden insbesondere in Großforschungsprojekten mittels Teleskopen, Supercomputern oder Teilchenbeschleunigern produziert. In diesem Rahmen wurde sogar ein viertes Paradigma, nämlich das der datengetriebenen Wissenschaft ausgerufen.² Im Kielwasser dieser riesigen Datenflut schwimmen jede Menge dunkle Daten.³ Dies bezeichnet Daten, die nicht ohne Weiteres (nach-)nutzbar, d.h. für Dritte oder auch nach einer gewissen Zeit nicht ohne weiteren Aufwand zugreifbar und verstehbar sind. Sie sind kein reines Phänomen der Wissenschaft,⁴ im Bereich der kommerziellen Datenverarbeitung gehen Schätzungen davon aus, dass die Klasse der dunklen Daten bis zu 80 % ausmacht.^{5,6}

Im kommerziellen Bereich werden die bewusst und als Schutzmaßnahme im Dunkeln liegenden Datenspuren der Nutzer:innen (personenbezogene oder personenbeziehbare Daten, die das Verhalten im Internet abbilden) durch Datentracking zum Zwecke der Monetarisierung an das

1 Tony Hey und Anne Trefethen: »The Data Deluge: An e-Science Perspective«, in: Fran Berman, Geoffrey Fox und Tony Hey (Hg.): *Grid Computing: Making the Global Infrastructure a Reality*, Hoboken 2003, S. 809–824.

2 Tony Hey, Stewart Tansley und Kristin Tolle, K. M. (Hg.): *The Fourth Paradigm: Data-Intensive Scientific Discovery*, Washington 2009.

3 Björn Schembera und Juan M. Durán: »Dark Data as the New Challenge for Big Data Science and the Introduction of the Scientific Data Officer«, in: *Philosophy & Technology* 33 (2020), Heft 1, S. 93–115.

4 Im wissenschaftlichen Bereich sind dunkle Daten zumeist unerwünscht (außer beispielsweise bei medizinischen Daten), im kommerziellen Sektor können dunkle Daten für Einzelpersonen durchaus wünschenswert sein, um die Privatsphäre zu schützen.

5 Michael O'Reilly: »The Unseen Data Conundrum«, in: *Forbes*, 3.2.2022, <https://www.forbes.com/sites/forbestechcouncil/2022/02/03/the-unseen-data-conundrum/> (aufgerufen: 12.5.2023).

6 Die ›Data‹-Unterdisziplinen der Informatik wie Data Science, Data Analytics oder Data Engineering haben allesamt zum Ziel, Licht ins Dunkel der Daten zu bringen.