



Physikalische Chemie, Universität Rostock

**Vorlesung: Forschungsdatenmanagement
im Sommersemester 2024**

Dr. habil. Till Biskup

— Glossar zu Vorlesung 03: „Was ist Forschungsdatenmanagement?“ —

Hinweis: Die nachfolgend genannten Begriffe und Definitionen erheben keinen Anspruch auf formale Korrektheit, sondern dienen lediglich dem besseren Verständnis der in der Vorlesung behandelten Themen und sind im jeweiligen Kontext zu sehen. Mehrfache, voneinander abweichende Definitionen in unterschiedlichen Kontexten sind daher möglich. Fremdsprachige Begriffe werden nach Möglichkeit übersetzt, erscheinen aber ggf. unter ihrem ursprünglichen Namen in der Liste. Verweise auf andere Begriffe innerhalb des Glossars sind durch das vorangestellte Symbol ↑ gekennzeichnet.

Big Data nach Gartner Informationen mit großem Umfang, hoher Geschwindigkeit und/oder großer Vielfalt, deren Verarbeitung kosteneffiziente und innovative Werkzeuge erfordert, die sich positiv auf Einblicke, Entscheidungsfindungen und die Automatisierung von Prozessen auswirken. Ein Wesensmerkmal von Big Data ist die Verwendung der Daten in einem anderen Kontext als jenem, in dem sie ursprünglich erhoben wurden. Das führt meist zu einer ganzen Reihe von Problemen, die aber den Datennutzenden nicht unbedingt bewusst sind. Insbesondere ist der Kontext der Datenerhebung (fast) nie ausreichend dokumentiert (und dokumentierbar), um einer fachfremden Person die Einschätzung zu erlauben, ob die Verwendung der Daten im gegebenen Kontext zulässig ist.

Big Science siehe ↑Großforschung

Cargo Cult Phänomen, dass Indigene auf Südseeinseln, die im Zweiten Weltkrieg als US-amerikanische Stützpunkte dienten und über denen in dieser Zeit aus Flugzeugen große Mengen Nahrungsmittel und andere Güter abgeworfen wurden, nach dem Krieg, als die Lieferungen ausblieben, kultartig das Verhalten der Soldaten in der Hoffnung nachstellten, wieder Güter zu erhalten. So wurden Landebahnen und Flughafentower genauso imitiert

wie das Verhalten der Soldaten auf den vorigen Flugplätzen.

Cargo Cult Science von Richard Feynman [1] eingeführter Begriff für eine Form der (vermeintlichen) Wissenschaft, die allen offensichtlichen Vorzeichen und Formen wissenschaftlicher Forschung folgt, der aber etwas Wesentliches fehlt: wissenschaftliche Integrität, Ehrlichkeit, rigoroses Hinterfragen der eigenen Ergebnisse und Erklärungen und Offenlegung aller relevanten Informationen nach bestem Wissen und Gewissen. Vgl. ↑Cargo Cult

datengetriebene Wissenschaft „viertes Paradigma“, von Jim Gray [2] maßgeblich geprägter Begriff; beschreibt das Betreiben von Wissenschaft ausgehend von verfügbaren Daten. Die Fragestellung wird durch die Daten und deren Verfügbarkeit bestimmt, nicht umgekehrt. Nur möglich durch die unter dem Begriff ↑e-Science zusammengefassten Werkzeuge und Infrastrukturen.

Erkenntnis Aneignung des Sinngehalts von erlebten bzw. erfahrenen Sachverhalten, Zuständen oder Vorgängen, Ergebnis des Vorgangs des Erkennens. Erkenntnis beinhaltet immer eine auf die Erfahrung gestützte Beurteilung und setzt notwendiger Weise ein Subjekt voraus, das erkennt. Neue Erkenntnisse, die von innerer und äußerer Erfahrung unabhängig

sind, sind immer Ergebnis einer schöpferischen Phantasie. Bei der Erkenntnis stehen sich Subjekt und Objekt als Erkennendes und Erkanntes gegenüber. Die Erkenntnis führt zu einem Abbild des Objekts im Subjekt. Die grundsätzliche Unvollständigkeit dieses Abbilds ist die Triebkraft hinter dem Erkenntnisgewinn und letztlich der \uparrow Wissenschaft. Vgl. [3]; wesentliche Beiträge zur Erkenntnistheorie und ihrer Anwendung auf die Naturwissenschaft kommen von Kant [4, 5].

e-Science Summe der digitalen Werkzeuge und der notwendigen digitalen Infrastruktur, um mit großen Datenmengen umzugehen; Voraussetzung für die \uparrow datengetriebene Wissenschaft, aber von dieser unabhängig.

FAIR Akronym für die vier Begriffe *findable* (auffindbar), *accessible* (zugreifbar), *interoperable* (interoperabel) und *reusable* (wiederverwendbar); von Wilkinson *et al.* [6] unter dem vollständigen Titel „The FAIR Guiding Principles for scientific data management and stewardship“ berühmt gemachte Prinzipien, die aus der \uparrow datengetriebenen Wissenschaft und der Verwendung von \uparrow künstlicher Intelligenz zur Verarbeitung großer Datenmengen kommen. Oft missverstanden als tragfähiges Grundkonzept für Forschungsdatenmanagement. Für die meisten Forschenden in ihrer originalen Form eher irrelevant, aber für die Wissenschaft und den Erkenntnisgewinn tendenziell gefährlich.

Forschung Systematisches Vorgehen, um einer bestimmten Fragestellung nachzugehen oder Phänomene zu erklären oder Experimente unter kontrollierten Bedingungen durchzuführen, das der wissenschaftlichen Methodik folgt. Wissenschaft setzt Forschung voraus. Allerdings kann Forschung ohne Beitrag zur Wissenschaft (ohne Erkenntnisgewinn) bleiben, vgl. Feynmans Begriff \uparrow Cargo Cult Science [1].

Forschungsdaten zunächst einmal Daten, die im Zuge wissenschaftlicher Vorhaben im Rahmen von Forschung z.B. durch Digitalisierung, Quellenforschungen, Experimente, Mes-

sungen, Erhebungen oder Befragungen entstehen. Forschungsdaten im weiteren Sinn umfassen darüber hinaus (physische) Objekte und Werkzeuge (z.B. Fragebögen, Software und Simulationen). Forschungsdaten können grundsätzlich analog oder digital vorliegen. Sie sind Ausgangspunkt der (empirischen) Wissenschaft.

Forschungsdatenmanagement Umgang mit Forschungsdaten über ihren gesamten Lebenszyklus hinweg mit dem Fokus auf Nachvollziehbarkeit und Nachnutzbarkeit; wird meist auf die digitale Welt bezogen, ist letztlich aber nichts anderes als sauberes wissenschaftliches Arbeiten; notwendige, aber nicht hinreichende Bedingung für den wissenschaftlichen Erkenntnisgewinn.

Großforschung *Big Science*, außeruniversitär, meist in großen Forschungseinrichtungen teils quasi-industriell betriebene Form der Wissenschaft, die seit Mitte des 20. Jh. verstärkt auftritt. Meist wird der Beginn mit dem Manhattan-Projekt (Los Alamos National Laboratory) gleichgesetzt. Als Begriff von Alvin Weinberg [7] eingeführt und von Derek J. de Solla Price [8] berühmt gemacht. Nach dem zweiten Weltkrieg einsetzendes Phänomen, dass sich die Politik stark für die Wissenschaft interessierte und große Forschungsstandorte mit sehr viel Geld förderte. Die wissenschaftlichen Fragestellungen werden von den Großforschungsgeräten und -Einrichtungen bestimmt, nicht umgekehrt. Nach Weinberg besteht die Gefahr, dass das nun vorhandene Geld ausgegeben wird, anstatt erst sinnvoll nachzudenken.

Infrastruktur personelle, sachliche und finanzielle Ausstattung, um ein angestrebtes Ziel zu erreichen.

Konvention innerhalb einer Gruppe oder einem (lokalen) Kontext getroffene (temporäre) Festlegung. Ziel von Konventionen ist die Vereinheitlichung und damit einhergehend die Befreiung von der Notwendigkeit, jedesmal aufs Neue nachdenken zu müssen, wie z.B. gewisse Prozesse durchgeführt oder Objekte benannt

werden sollen. Konventionen sind im Gegensatz zu ↑Standards weniger verbindlich und deutlich flexibler sowie *ad hoc* innerhalb einer Gruppe einführbar. Vgl. ↑Standard

künstliche Intelligenz (KI), meist besser beschrieben als „maschinelles Lernen“ (ML); aktuell wieder einmal sehr populär und als Heilsversprechen gehandelt. Letztlich in seiner momentanen Ausprägung die Anwendung (komplexerer) statistischer Algorithmen auf große Datenmengen.

Management (Tätigkeit) Planung, Organisation und Verwaltung (und ggf. Leitung) unter ökonomischen Gesichtspunkten. Auch wenn im Kontext der Wissenschaft ökonomische Aspekte oft nur eine untergeordnete Rolle spielen, ist ein ökonomischer Ressourceneinsatz kein Widerspruch zur (und keine Einschränkung der) wissenschaftlichen Vorgehensweise.

Metadaten Informationen zu den numerischen Daten, notwendige Voraussetzung für eine sinnvolle Verarbeitung der Daten im Kontext eines ↑Systems zur Datenverarbeitung und für ↑nachvollziehbare Wissenschaft.

nachvollziehbare Wissenschaft *reproducible science*, seit der Etablierung rechnergestützter Datenauswertung eigentlich nie mehr erreicht, aber für die Wissenschaft konstituierender Aspekt, dass sich Ergebnisse und Auswertungen unabhängig nachvollziehen lassen, weil alle dazu notwendigen Aspekte vollständig und ausreichend beschrieben wurden (↑Nachvollziehbarkeit). Motivation für die Vorlesung, deren Ziel es ist, die Hörer mit Konzepten vertraut zu machen, die letztlich eine ernstzunehmende nachvollziehbare Wissenschaft ermöglichen. Die ↑Nachvollziehbarkeit geht dabei weit über ↑Replizierbarkeit und ↑Reproduzierbarkeit hinaus.

Nachvollziehbarkeit zentraler Aspekt der ↑Wissenschaft und der wissenschaftlichen Methode, die die Intersubjektivität ihrer Aussagen ermöglicht. Setzt in der Regel eine ↑hinreichende Beschreibung (und Dokumentation) der einzelnen Schritte voraus, die von einem

gegebenen Ausgangspunkt zu einem (neuen) Ergebnis oder auch einer Erkenntnis kommt.

Replizierbarkeit *replicability*, unabhängige Wiederholbarkeit der (Roh-)Datenerhebung, meist in Form von Experimenten und Beobachtungen, entsprechend nicht in jedem Fall durchführbar. Vgl. ↑Reproduzierbarkeit, ↑Robustheit, ↑Verallgemeinerbarkeit.

Reproduzierbarkeit *reproducibility*, vollständige Wiederholbarkeit einer beschriebenen Datenverarbeitung und -Analyse. Ausgangspunkt sind existierende Daten, entsprechend sollte sie in jedem Fall möglich sein. Vgl. ↑Replizierbarkeit.

Robustheit *robustness*, im Kontext der Datenverarbeitung die Tatsache, dass unterschiedliche, unabhängige Analysen derselben Daten zum gleichen Ergebnis führen. Vgl. ↑Reproduzierbarkeit, ↑Replizierbarkeit, ↑Verallgemeinerbarkeit

Standard von einem oft internationalen und anerkannten Gremium definierte Festlegung. Standards sind im Gegensatz zu ↑Konvention sehr viel starrer und nicht *ad hoc* von einer Gruppe einführbar. Vgl. ↑Konvention

System zur Datenverarbeitung hier: Gesamtsystem für wissenschaftliche Datenverarbeitung von der Datenaufnahme bis zur fertigen Publikation, das alle Aspekte umfasst und das ↑nachvollziehbare Wissenschaft möglich macht und gewährleistet. Definitiv ein größeres Projekt, das nicht nur eine ↑monolithische Anwendung umfasst, sondern viele Aspekte darüber hinaus. Setzt entsprechende ↑Infrastruktur und in der Umsetzung der einzelnen Komponenten sauberen Code und eine solide Softwarearchitektur voraus.

Verallgemeinerbarkeit auch: Generalisierbarkeit, *generalisability*, im Kontext der Datenverarbeitung die Tatsache, dass sowohl unabhängig erhobene Daten als auch voneinander unabhängige Analysemethoden zum gleichen Ergebnis führen. Baustein zur unabhängigen Bestätigung wissenschaftlicher Hypothesen. Vgl. ↑Reproduzierbarkeit, ↑Replizierbarkeit, ↑Robustheit

viertes Paradigma *fourth paradigm*, siehe ↑datengetriebene Wissenschaft

Wissenschaft Auf den Erkenntnisgewinn ausgerichtete, systematisches menschliches Unterfangen, das in der Regel eine Reihe von Kri-

terien erfüllt bzw. erfüllen sollte: Unabhängigkeit vom Beobachtenden bzw. Durchführenden, gegründet auf den Erkenntnissen früherer Generationen, sowie überprüfbar, nachvollziehbar und ggf. reproduzierbar. Für Einführungen vgl. u.a. [9, 10].

Literatur

- [1] Richard P. Feynman. Cargo cult science. 37.7 (1974), S. 10–13. URL: <https://resolver.caltech.edu/CaltechES:37.7.CargoCult>.
- [2] Tony Hey, Stewart Tansley und Kristin Tolle, Hrsg. *The Fourth Paradigm*. Redmont, Washington: Microsoft Research, 2009.
- [3] Heinrich Schmidt. *Philosophisches Wörterbuch*. 22. Aufl. Neu bearbeitet von Georgi Schischkoff. Stuttgart: Kröner, 1991.
- [4] Immanuel Kant. *Kritik der reinen Vernunft*. Herausgegeben von Wilhelm Weischedel. Frankfurt am Main: Suhrkamp, 1974.
- [5] Immanuel Kant. *Metaphysische Anfangsgründe der Naturwissenschaft*. Mit einer Einleitung herausgegeben von Konstantin Pollok. Hamburg: Felix Meiner Verlag, 1997.
- [6] Mark D. Wilkinson u. a. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3 (2016), S. 160018. DOI: 10.1038/sdata.2016.18.
- [7] Alvin Weinberg. Impact of large-scale science on the United States. *Science* 134 (1961), S. 161–164. DOI: 10.1126/science.134.3473.161.
- [8] Derek John Price de Solla. *Little science, big science*. New York: Columbia University Press, 1963.
- [9] Alan F. Chalmers. *What is this thing called Science?* Third edition. Berkshire, UK: Open University Press, 1999.
- [10] Hans Poser. *Wissenschaftstheorie*. Stuttgart: Reclam, 2001.