



Physikalische Chemie, Universität Rostock

**Vorlesung: Forschungsdatenmanagement
im Sommersemester 2024**

Dr. habil. Till Biskup

— Glossar zu Vorlesung 07: „Auswerten“ —

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.

Abstraktion Nach Edsger Dijkstra [1] das einzige mentale Werkzeug, das es erlaubt, eine große Vielzahl von Fällen abzudecken. Zweck der Abstraktion ist es nicht, vage zu sein, sondern im Gegenteil ein neues Bedeutungsniveau zu schaffen, das präzise Beschreibungen erlaubt.

Automatisierung *automation* Strategie, um sich die manuelle Durchführung repetitiver und meist langweiliger Prozesse zu ersparen, indem sie an Maschinen ausgelagert wird. Automatisierung sorgt für Konsistenz (aber nicht Fehlerfreiheit) und ermöglicht es den menschlichen Akteuren, ihre dadurch freiwerdende Kapazität auf die eigentlichen intellektuellen Aufgaben, die weder automatisiert noch von Algorithmen übernommen werden können, zu verwenden.

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 ↑Wissenschaft. Vgl. [2]; wesentliche Beiträge zur Erkenntnistheorie und ihrer Anwendung auf die Naturwissenschaft kommen von Kant [3, 4].

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

freie Software Software, die Endnutzern die

Freiheiten der Nutzung, des Überprüfens/Studierens, des Teilens und des Modifizierens der Software gewährt. Das erfordert zwingend die Offenlegung des Quellcodes der Software. Der Begriff „frei“ bezieht sich hier auf die Freiheiten des Nutzers, nicht auf die Kosten der Software. Die Begriffe „*open source*“ und „freie Software“ sind nicht gänzlich deckungsgleich.

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

Konsistenz hier: logische Widerspruchsfreiheit; Zusammenhang der Gedankenführung

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

Lizenz *license*, Nutzungsrecht; u.a. Software ist *per se* vom Urheberrecht geschützt, unabhängig von ihrer Funktionalität. Lizenzen übertragen Nutzungsrechte vom Urheber der Software an ihren Nutzer. Inwieweit ↑Forschungsdaten urheberrechtlich geschützt sind, ist eine in der Rechtsprechung noch nicht abschließend beantwortete Frage. Tendenziell sind Daten, die nicht weiter kuratiert wurden, nicht urheberrechtlich geschützt, da ihnen die nötige Schöpfungshöhe fehlt.

Medienbruch Übertragung von Daten bzw. Informationen von einem auf ein anderes Daten- bzw. Informationsmedium. Beispiele sind das manuelle Abtippen von Informationen, aber auch der Wechsel von Programmen oder Rechnern. Medienbrüche verlangsamen und verschlechtern Datenverarbeitungsprozesse und können zu Übertragungsfehlern führen. Selbst wenn Daten bzw. Informationen zwischen Prozessen in digitalen Forma-

ten ausgetauscht werden können, ist ein Informationsverlust nur dann gewährleistet, wenn das Datenmodell des empfangenden Prozesses mindestens gleich mächtig ist wie das des sendenden Prozesses und es eine (eindeutige) Abbildung des einen auf das andere Datenmodell gibt.

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

Modell vereinfachende Annäherung an die Wirklichkeit bzw. einen Forschungsgegenstand, die nur die als wesentlich erachteten Phänomene als Aspekte berücksichtigt. Ein Modell entsteht oft durch ↑Abstraktion und Verallgemeinerung von Beobachtungen. Mathematisch formulierte Modelle zeichnen sich häufig durch Parameter aus, durch die sie sich charakterisieren lassen und die ggf. variiert werden können, um das Modell an die Realität bzw. erhobene Daten anzupassen. Während Modelle der einzige Weg sind, die Realität zu erklären und so ggf. zu ↑Erkenntnis zu gelangen, sind sie immer vorläufig und beschränkt. Von George E. P. Box stammt der Satz „Alle Modelle sind falsch, aber manche sind nützlich“ [5].

Modularisierung Aufteilung der Gesamtaufgabe in kleinere Abschnitte. Die Aufteilung wird so lange fortgesetzt, bis die Lösung für den aktuellen Abschnitt unmittelbar offensichtlich ist.

nachvollziehbare Wissenschaft *reproducible science*, seit der Etablierung rechnergestützter Datenauswertung eigentlich nie mehr erreichter, 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 ↑Nachvollziehbar-

keit 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.

Persistenz Fähigkeit, Daten oder logische Verbindungen über lange Zeit (insbesondere über einen Programmabbruch hinaus) bereitzuhalten; benötigt ein nichtflüchtiges Speichermedium.

PID *persistent identifier*, dt. dauerhafte Kennung, i.d.R. eindeutige und langzeitstabile Kennung für physische oder digitale Objekte. Bekannte und weit verbreitete PIDs sind z.B. der *digital object identifier* (DOI), aber auch die *International Standard Book Number* (ISBN) oder die *Open Researcher and Contributor ID* (ORCID). Vgl. ↑Persistenz

Plausibilität (kontextabhängiges) Beurteilungskriterium: etwas ist plausibel, wenn es möglich und wahrscheinlich erscheint.

Qualitätskontrolle Überprüfung der Qualität von Dingen oder Prozessen anhand vorher festgelegter Kriterien. Allgemeine Kriterien sind ↑Konsistenz und ↑Plausibilität. Wenn sich die Kriterien formal definieren und die relevanten Charakteristika der zu überprüfenden Dinge oder Prozesse ohne direkte menschliche Interaktion bestimmen lassen, ist eine Automatisierung möglich. Vgl. ↑Qualitätssicherung

Qualitätssicherung Sicherstellung der Qualität von Dingen oder Prozessen. Vgl. ↑Qualitätskontrolle

quelloffen Software, deren Quellcode offengelegt ist, so dass jeder ihn einsehen kann. Quelloffenheit ist eine notwendige, aber keine hinreichende Voraussetzung für ↑freie Software.

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.

Revision einzelner der ↑Versionsverwaltung bekannter Zustand (↑Version)

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

Softwareentwicklung umfassender systematischer Prozess, der alle Lebensstadien einer Software von der Planung über die Implementierung bis hin zur Pflege und ggf. der Abwicklung (und dem Ersatz durch Nachfolger) umfasst. Nicht zu verwechseln mit Programmierung, die bestenfalls einen Teil der Implementierung und Pflege von Software umfasst. Wird im wissenschaftlichen Kontext meist mindestens unterschätzt, wenn nicht offen ignoriert, mit entsprechend verheerenden Folgen für die ↑Nachvollziehbarkeit und damit Wissenschaftlichkeit der auf Software basierender oder von ihr unterstützter Forschung.

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

Version eindeutiger Zustand einer Software oder eines Dokuments. Zur ↑Nachvollziehbarkeit bedarf es einer strukturierten Versionierung mit Hilfe einer ↑Versionsverwaltung und zum Verweis auf eine Version typischerweise einer ↑Versionsnummer.

Versionsnummer hier: eindeutige Bezeichnung einer ↑Version einer Software oder einer Pro-

zessbeschreibung, deren Kenntnis es erlaubt, auf genau diese Version der Software/Prozessbeschreibung Bezug zu nehmen.

Versionsverwaltung *version control system, VCS*; Software zur Verwaltung unterschiedlicher ↑Versionen von Dateien und Programmen, die den Zugriff auf beliebige ältere als Versionen (↑Revision) gespeicherte Zustände ermöglicht. Gleichzeitig ein wichtiges Werkzeug für die Softwareentwicklung und wesentlicher Aspekt einer Projektinfrastruktur.

Wissenschaft Auf den Erkenntnisgewinn ausgerichteter, systematisches menschliches Unterfangen, das in der Regel eine Reihe von Kriterien 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. [6, 7].

Literatur

- [1] Edsger W. Dijkstra. The humble programmer. *Communications of the ACM* 15 (1972), S. 859–865.
- [2] Heinrich Schmidt. *Philosophisches Wörterbuch*. 22. Aufl. Neu bearbeitet von Georgi Schischkoff. Stuttgart: Kröner, 1991.
- [3] Immanuel Kant. *Kritik der reinen Vernunft*. Herausgegeben von Wilhelm Weischedel. Frankfurt am Main: Suhrkamp, 1974.
- [4] Immanuel Kant. *Metaphysische Anfangsgründe der Naturwissenschaft*. Mit einer Einleitung herausgegeben von Konstantin Pollok. Hamburg: Felix Meiner Verlag, 1997.
- [5] G.E.P. Box. „Robustness in the Strategy of Scientific Model Building“. In: *Robustness in Statistics*. Hrsg. von Robert L. Launer und Graham N. Wilkinson. Academic Press, 1979, S. 201–236. ISBN: 978-0-12-438150-6. DOI: <https://doi.org/10.1016/B978-0-12-438150-6.50018-2>.
- [6] Alan F. Chalmers. *What is this thing called Science?* Third edition. Berkshire, UK: Open University Press, 1999.
- [7] Hans Poser. *Wissenschaftstheorie*. Stuttgart: Reclam, 2001.