



Physikalische Chemie, Universität Rostock

**Vorlesung: Forschungsdatenmanagement
im Sommersemester 2024**

Dr. habil. Till Biskup

— Glossar zu Vorlesung 09: „Veröffentlichen“ —

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.

Datenformat digitales Speicherformat für Daten jeglicher Form. Grundsätzlich werden binäre und Textformate unterschieden. Während erstere meist mit deutlich geringerem Speicherbedarf auskommen, sind sie im Gegensatz zu letzteren nicht ohne Hilfsmittel lesbar. Textformate hingegen sind, ein beliebiger Texteditor vorausgesetzt, prinzipiell menschenlesbar.

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

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 Nutzbarkeit; 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.

hinreichend mathematisches Konzept, das eine Bedingung beschreibt, deren Erfüllung ausreicht, um ein gegebenes Ziel zu erreichen. Vgl. ↑notwendig

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

Forschungsdaten zunächst einmal Daten, die im

Katalog Werkzeug zum Auffinden und Erschließen

von Forschungsdaten. ↑Forschungsdaten können mit Hilfe eines Datenkatalogs gesucht, gefunden und erschlossen werden. Ein Datenkatalog enthält vergleichbar zu einem Bibliothekskatalog verschiedene ↑Metadaten, die die Grundlage für die Suche und Filterung darstellen, aber nicht (notwendigerweise) die ↑Forschungsdaten selbst – im Falle der Bibliothek die Bücher. Typischerweise bieten auch ↑Repositorien grundständige Katalogfunktionen, so dass die Unterscheidung zwischen Katalog und Repository in der Praxis mitunter verschwimmt. Ein Katalog als Sammlung von ↑Metadaten zu bestimmten Objekten erweist sich insbesondere dann als sinnvoll, wenn die Menge der Objekte eine gewisse Schwelle überschreitet, die ein Auffinden und Abrufen über die einzelnen Objekte selbst unmöglich macht oder zumindest massiv erschwert.

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.

Metadaten Informationen zu den numerischen Daten, notwendige Voraussetzung für eine sinn-

volle 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 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 ↑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.

notwendig mathematisches Konzept, das eine Bedingung beschreibt, die zwar erfüllt sein muss, um ein bestimmtes Ergebnis zu bekommen, aber für die Erfüllung nicht ausreicht. Vgl. ↑hinreichend

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.

Präregistrierung insbesondere bei Studien mit Tieren und Menschen zunehmend verfolgtes Konzept, dass die Studien vorab bei einem (nationalen) Gremium hinterlegt werden müssen. Dient dazu, die Ergebnisse dieser Studien in jedem Fall zu veröffentlichen, auch wenn sie negativ ausfallen sollten, um größtmöglichen Nutzen für alle daraus zu ziehen. In Deutschland ist das Deutsche Zentrum zum Schutz von Versuchstieren (Bf3R) Bestandteil des Bundesinstituts für Risikobewertung (BfR).

proprietär auf herstellerspezifischen, (meist) nicht veröffentlichten Standards basierend

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

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

Repository Publikationsplattform (u.a.) für ↑Forschungsdaten. Repositorien sind Publikationsplattformen (u.a.) für Forschungsdaten. Als IT-Dienst werden sie i.d.R. von Institutionen, Organisationen oder Firmen bereitgestellt und speichern die Forschungsdaten i.d.R. langfristig, dokumentieren die Forschungsdaten mit ↑Metadaten, regeln den Zugang (inkl. ↑Lizenz) zu den Forschungsdaten und vergeben einen ↑PID. Die dort publizierten Forschungsdaten sind meist über eine

Metadatensuche und -filterung für Nutzerinnen und Nutzer auffindbar und erschließbar (Datenkatalog). Vgl. ↑Katalog

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.

Transparenz über die ↑Nachvollziehbarkeit hinausgehendes Konzept, das die Wege der Entscheidungsfindung inklusive verworfener oder nicht beschrittener Alternativen nach bestem Wissen und Gewissen umfassend dokumentiert. Von R. Feynman [4] als essentiell für die Wissenschaftlichkeit hervorgehoben.

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 Prozessbeschreibung, 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 ausgerichtete, 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. [5, 6].

Literatur

[1] Heinrich Schmidt. *Philosophisches Wörterbuch*. 22. Aufl. Neu bearbeitet von Georgi Schischkoff. Stuttgart: Kröner, 1991.

[2] Immanuel Kant. *Kritik der reinen Vernunft*. Herausgegeben von Wilhelm Weischedel. Frankfurt am Main: Suhrkamp, 1974.

[3] Immanuel Kant. *Metaphysische Anfangsgründe der Naturwissenschaft*. Mit einer Einleitung herausgege-

ben von Konstantin Pollok. Hamburg: Felix Meiner Verlag, 1997.

[4] Richard P. Feynman. Cargo cult science. 37.7 (1974), S. 10–13. URL: <https://resolver.caltech.edu/CaltechES:37.7.CargoCult>.

[5] Alan F. Chalmers. *What is this thing called Science?* Third edition. Berkshire, UK: Open University Press, 1999.

[6] Hans Poser. *Wissenschaftstheorie*. Stuttgart: Reclam, 2001.