



Physikalische Chemie, Universität Rostock

**Vorlesung: Forschungsdatenmanagement
im Sommersemester 2024**

Dr. habil. Till Biskup

— Glossar zu Vorlesung 05: „Planen“ —

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.

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 [4] 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

Datenmanagementplan DMP, *data management plan* (1.) von Fördermittelgebern mitunter gefordertes formales Dokument, das den Umgang mit in einem Projekt erhobenen ↑Forschungsdaten inklusive Nachnutzung beschreibt; (2.) durch akzeptierte interne Konventionen und Strukturen geprägter bewuss-

ter Umgang mit ↑Forschungsdaten mit dem Fokus auf ↑Nachvollziehbarkeit und Nutzbarkeit

DMP ↑Datenmanagementplan

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].

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 vor-

aus. Allerdings kann Forschung ohne Beitrag zur Wissenschaft (ohne Erkenntnisgewinn) bleiben, vgl. Feynmans Begriff ↑Cargo Cult Science [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 (↑Erkenntnis).

Forschungsdatenlebenszyklus Modell der wissenschaftlichen Methode aus Sicht der ↑Forschungsdaten. Der Forschungsdatenlebenszyklus ist ein für Forschende intuitives Modell, das es ermöglicht, mit ihnen über das ↑Forschungsdatenmanagement (und damit Dinge, die sie direkt betreffen) ins konstruktive Gespräch zu kommen. Der Forschungsdatenlebenszyklus hilft dabei, alle wesentlichen Aspekte in den Blick zu nehmen bzw. im Blick zu behalten. Damit ist er vergleichbar z.B. entsprechenden Entwicklungszyklen für Software oder allgemein Projekte. Letztlich ist er damit ein Planungsinstrument. Der Forschungsdatenlebenszyklus ist lediglich ein Modell, und es gibt vermutlich fast so viele Varianten wie Menschen, die sich damit auseinandergesetzt haben. Der Forschungsdatenlebenszyklus ist ein Zyklus, aber das bedeutet nicht zwangsläufig, dass alle Stufen nacheinander abgearbeitet werden. Dazu kommt, dass je nach Forschungsansatz unterschied-

liche Aspekte unterschiedlich gewichtet werden. Der Zykluscharakter bedeutet, dass es sich um iterative (und letztlich rückgekoppelte) Prozesse handelt.

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.

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 sinnvolle Verarbeitung der Daten im Kontext eines ↑Systems zur Datenverarbeitung und für ↑nachvollziehbare Wissenschaft.

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

monolithisch aus einem Stück bestehend; zusammenhängend und fugenlos

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

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

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

Wissenschaft Auf den Erkenntnisgewinn (↑Erkenntnis) ausgerichtetes, 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.