
Forschung und Wissenstransfer

Leitlinie zum Umgang mit Forschungsdaten

Impressum

TH Köln
Gustav-Heinemann-Ufer 54
50968 Köln

Herausgeber:
Präsidium der TH Köln

Text: Katharina Sofie Sauther-Patrascu
Mitwirkung für IT: Sofia Dmitrieva

Bei Fragen zur Leitlinie und zum Forschungsdatenmanagement wenden Sie sich an: forschungsdaten@th-koeln.de.

Stand: 11/2021

1 Präambel

Die Technische Hochschule Köln (TH Köln) steht für eine gelebte Qualitätskultur. Diese bestimmt unseren Hochschulalltag in allgemeinen Prozessen der Bildung, der Forschung und des Transfers sowie des Managements und der Services. Im Forschungskontext ist daher ein verantwortungsvoller Umgang mit *Forschungsdaten*¹ für diesen Qualitätsanspruch essenziell. Ein solcher Umgang mit Forschungsdaten fördert die Transparenz, die wissenschaftliche Überprüfbarkeit und die Nachnutzbarkeit von Forschungsergebnissen.

Aus dieser Leitlinie ergibt sich der Rahmen zum verantwortungsvollen Management von Forschungsdaten an der TH Köln.² Im Einklang mit der Open-Access-Erklärung³ sowie der Ordnung zur Sicherung guter wissenschaftlicher Praxis⁴ richtet sich diese Richtlinie an alle forschenden Mitglieder und Angehörige der TH Köln. Angesprochen sind *Forschende* in ihrer Rolle als eigenständige Wissenschaftlerinnen und Wissenschaftler oder als wissenschaftlicher Nachwuchs sowie die unterstützenden Serviceeinheiten. Aber auch Lehrende stehen hier in der Verantwortung in der Rolle als Betreuende und Vorbilder für Studierende und den wissenschaftlichen Nachwuchs, für die diese Leitlinie gleichermaßen Gültigkeit besitzt.

2 Grundsätze

- (1) Der Begriff *Forschungsdaten* umfasst in dieser Leitlinie Daten, die während des Forschungsprozesses generiert, erhoben oder genutzt werden oder dessen Ergebnis sind, unabhängig von der Art der Entstehung oder des Speicherformats. Die Vielfalt der Forschungsdaten spiegelt somit die Vielfalt der wissenschaftlichen Disziplinen, der Erkenntnisinteressen und der Forschungsmethoden wider.
- (2) Das *Forschungsdatenmanagement* betrifft den gesamten Lebenszyklus von Forschungsdaten und die Planung der einzelnen Schritte im Forschungsprozess von der Erfassung, Dokumentation, Verarbeitung über die Aufbewahrung bis zur Veröffentlichung bzw. *Archivierung* oder auch Löschung von Forschungsdaten.
- (3) Forschende der TH Köln sollen mit ihren Daten *FAIR* umgehen, d.h. die Forschungsdaten sollen **F**indable (Auffindbar), **A**ccessible (Zugänglich), **I**nteroperable (Interoperabel) und **R**eusable (Wiederverwendbar) sein. Die *FAIR-Prinzipien* haben zum Ziel, dass Forschungsdaten für Menschen und Maschinen (z.B. Computern) optimal aufbereitet und zugänglich sind und damit die Auffindbarkeit von Datensätzen und deren Wiederverwendbarkeit gewährleistet ist.
- (4) Die TH Köln empfiehlt ausdrücklich die Erstellung eines *Datenmanagementplans* (DMP) vor oder zu Beginn eines jeden Forschungsvorhabens. Grundsätzlich wird den Forschenden angeraten, sich frühzeitig mit den Anforderungen, Kosten und Möglichkeiten der *Datenspeicherung*, Sicherung, Archivierung und ggf. Publikation von Forschungsdaten auseinanderzusetzen. Ein Datenmanagementplan regelt Authentizität, Integrität, Vollständigkeit, Auffindbarkeit, Vertraulichkeit und Veröffentlichung von Daten unter

¹ Alle kursiven Begriffe finden sich mit einer Definition im zur Leitlinie gehörigen Glossar.

² Mit unserer Leitlinie orientieren wir uns an der Leitlinie zum Umgang mit Forschungsdaten der DFG, Bonn 2015. https://www.dfg.de/download/pdf/foerderung/grundlagen_dfg_foerderung/forschungsdaten/leitlinien_forschungsdaten.pdf (25.01.2022)

³ Open-Access-Erklärung der Fachhochschule Köln. Verabschiedet vom Senat in seiner Sitzung vom 16.01.2013. Quelle: https://www.th-koeln.de/mam/downloads/deutsch/forschung/thk_open_access_erklarung.pdf (25.01.2022)

⁴ Ordnung zur Sicherung guter wissenschaftlicher Praxis der Technischen Hochschule Köln. TH Köln 2019. Quelle: https://www.th-koeln.de/mam/downloads/deutsch/hochschule/amtlicheMitteilungen/fassung_mit_anmerkungen_und_fussnoten.pdf (25.01.2022)

- Berücksichtigung der fachspezifischen Besonderheiten. In einigen Förderprogrammen ist die Erstellung eines Datenmanagementplans zwingend erforderlich.
- (5) Gute wissenschaftliche Praxis erfordert strenge Sorgfalt bei der Gewinnung und Auswertung von Daten und beinhaltet die eindeutige und nachvollziehbare Dokumentation des Entstehungskontextes und der benutzten Werkzeuge und Verfahren nach fachspezifischen Standards. Zur Dokumentation von Forschungsdaten gehört auch die Angabe von Quellen, Inhaberschaft und Berechtigung zur Nutzung und Veröffentlichung.
 - (6) *Primär-* und mit ihnen verbunden *Metadaten*, die als Grundlage für Veröffentlichungen dienen, müssen auf haltbaren und gesicherten Trägern für zehn Jahre (nach Projektende) aufbewahrt werden, um den Zugriff auf diese und die Nachvollziehbarkeit der Forschungsergebnisse zu gewährleisten.⁵
 - (7) Die TH Köln befürwortet Publikation von Forschungsdaten als Open Access im Sinne eines freien Zugangs zu Wissenschaft und Forschung, sofern keine rechtlichen oder ethischen Einschränkungen bestehen. Sofern die Forschungsdaten nicht gesondert veröffentlicht werden, muss ein Zugriff ermöglicht werden. Dafür müssen Forschungsdaten in einem geeigneten *Repository* oder Archivierungssystem abgelegt und auf eine korrekte, vollständige, unverfälschte und verlässliche Art und Weise gespeichert werden. Sofern eine Speicherung auf externen Systemen erfolgt, wird die Übermittlung der bibliografischen Daten in die Digitale Hochschulbibliografie⁶ empfohlen.
 - (8) Die Forschungsdaten sollen mit *persistenten Identifikatoren* (PID) versehen werden. Ein PID stellt sicher, dass Daten adressierbar und identifizierbar sind. Er kann während des Forschungsprozesses und zur Publikation von Forschungsdaten verwendet werden.
 - (9) Die TH Köln und ihre Forschenden beachten beim Forschungsdatenmanagement ethische Belange sowie rechtliche Vorgaben zu Datenschutz, Urheberrecht und Geheimhaltung oder von Dritten. Die Prüfung der Forschungsdaten im Sinne des Arbeitnehmererfindungsgesetzes bleibt hiervon unberührt.
 - (10) Fachspezifische Anforderungen z.B. in Bezug auf Metadaten sowie Vorgaben von Fördermittelgebern an das Forschungsdatenmanagement sind zu beachten.

3 Verantwortlichkeiten

- (1) Die Verantwortung für das Forschungsdatenmanagement liegt bei den Projektleiterinnen und Projektleitern sowie den Forschenden. Sie beginnt mit Erzeugung der Daten und endet mit deren endgültiger Löschung. Die Verantwortlichen bestimmen, zu welchem Zeitpunkt und zu welchen rechtlichen Bedingungen Forschungsdaten zugänglich gemacht werden.
- (2) Verlassen Forschende die Hochschule, muss der Verbleib der Forschungsdaten an der TH Köln entsprechend den rechtlichen Rahmenbedingungen mit der/dem Vorgesetzten geregelt sowie eine Kontaktperson benannt werden, die die Auffindbarkeit der Daten und den Zugriff auf diese gewährleistet.

⁵ Ordnung zur Sicherung guter wissenschaftlicher Praxis der Technischen Hochschule Köln. TH Köln 2019.
Quelle: https://www.th-koeln.de/mam/downloads/deutsch/hochschule/amtlichemitteilungen/fassung_mit_anmerkungen_und_fussnoten.pdf (25.01.2022)

⁶ https://www.th-koeln.de/hochschulbibliothek/digitale-hochschulbibliografie_82523.php (25.01.2022)

- (3) Die Prinzipien eines verantwortungsvollen Forschungsdatenmanagements müssen in den Unterweisungen zur guten wissenschaftlichen Praxis thematisiert werden.⁷

4 Beratung und Infrastruktur

- (1) Forschende erhalten beratende Unterstützung zu rechtlichen Belangen wie der Inhaberschaft von Forschungsdaten und deren Nutzungsrechte oder der Erstellung einer informierten Einwilligungserklärung zur Erhebung und Verarbeitung personenbezogener Interviewdaten für Forschungszwecke.⁸
- (2) Bei einer Übertragung von Nachnutzungs- und Veröffentlichungsrechten sollen in der Regel die Daten für wissenschaftliche Zwecke frei verfügbar bleiben.
- (3) Die TH Köln stellt geeignete Speichermöglichkeiten für Forschungsdaten sowie Plattformen für die inter- und transdisziplinäre Zusammenarbeit bereit.
- (4) Forschungsdaten werden auf Datenspeichern der TH Köln abgelegt oder bei Diensten, mit denen die TH Köln kooperiert.
- (5) Die TH Köln trägt dafür Sorge, dass ihre Mitglieder bei der Umsetzung des Forschungsdatenmanagements unterstützt werden. Dafür werden in Kooperation mit den Fakultäten Service- und Beratungsangebote entwickelt, die auch auf externe Dienstleister zurückgreifen können.

5 Gültigkeit und Umsetzung

Die Leitlinie zum Umgang mit Forschungsdaten an der Technischen Hochschule tritt am 10.11.2021 in Kraft und wird in den Amtlichen Mitteilungen der TH Köln veröffentlicht.

Ausgefertigt aufgrund des Beschlusses des Präsidiums der TH Köln vom 10.11.2021.

⁷ Ordnung zur Sicherung guter wissenschaftlicher Praxis der Technischen Hochschule Köln. TH Köln 2019.

Quelle:

https://www.th-koeln.de/mam/downloads/deutsch/hochschule/amtlichemitteilungen/fassung_mit_anmerkungen_und_fussnoten.pdf (25.01.2022)

⁸ Verantwortlich sind hier die Datenschutzbeauftragten des Justiziariats der TH Köln.

Glossar

Archivierung

Die *Archivierung* von Daten bedeutet sowohl die systematische Sicherung ausgewählter Daten für einen unbegrenzten Zeitraum als auch der Erhalt und die Betreuung dieser Daten über geeignete Verfahren. Im Gegensatz zum *Backup*, werden nur ausgewählte Daten in ihrer letzten Version archiviert. Eine erweiterte Form der *Archivierung* für mehr als 10 Jahre ist die *Langzeitarchivierung*.

Datenmanagementplan

Ein *Datenmanagementplan* (DMP) ist ein strukturierter Leitfaden zum Umgang mit *Forschungsdaten* während eines Forschungsvorhabens und über dessen Ende hinaus. Er dokumentiert als lebendiges Dokument den Entstehungsprozess von *Forschungsdaten* und wie sie ordnungsgemäß gespeichert werden, so dass sie in späteren Jahren interpretierbar und nachvollziehbar, verfügbar, authentisch, zitierfähig und nachnutzbar sind. Dazu sollen auch klar definierte rechtliche Parameter und geeignete Sicherheitsmaßnahmen (wie Verträge, Lizenzen) für die spätere Verwendung im DMP festgehalten werden.

Datenspeicherung/Backup

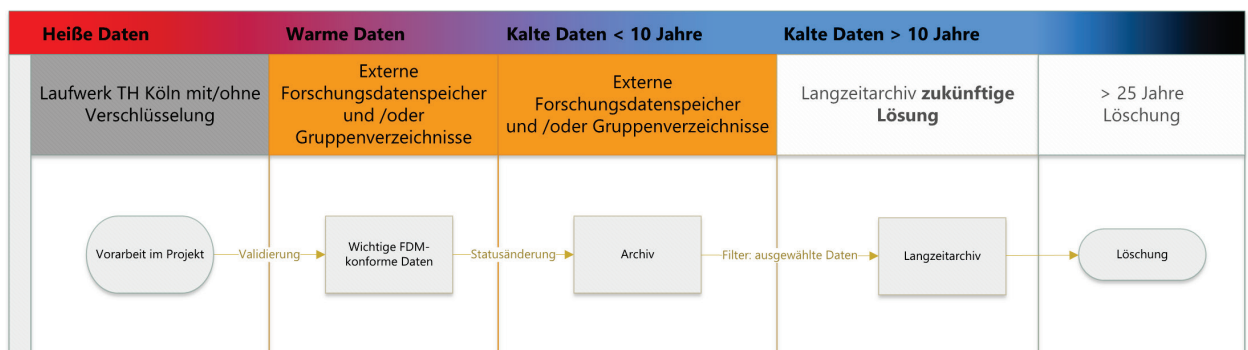
Die Sicherung von Daten wird meistens als ein Backup oder eine Sicherungskopie bezeichnet und dient der Wiederherstellung der ursprünglichen Daten im Falle eines Datenverlusts. Im Gegensatz zur *Archivierung*, bei der nur ausgewählte Daten gesichert werden, werden bei einem Backup oder einer Datensicherung alle für ein Projekt relevanten Daten abgespeichert.

Datenspeicherung und Archivierung an der TH Köln

Die Daten werden nach Art ihrer Zugriffshäufigkeiten in heiße, warme und kalte Daten mit einem unterschiedlichen Anspruch an die Speicherung untergliedert. Gleichzeitig bestehen zwischen diesen Arten fließende Übergänge und nicht immer klar zu definierende Grenzen.

- Heiße Daten: sehr häufiger Zugriff, aktive Projektarbeit, ggfs. zahlreiche Kopien/Versionen
- Warme Daten: Regelmäßiger Zugriff, jedoch stabile Datenzusammensetzung
- Kalte Daten: seltener Zugriff, ausgewählte archivierte Daten

Die folgende Abbildung dient als Orientierung, welche Datenarten wo abgespeichert werden.



Repositorium

Ein *Repositorium* ist eine Speicherplattform zur *Archivierung* und weltweiten Veröffentlichung von wissenschaftlichen Publikationen, *Forschungsdaten* oder Daten des kulturellen Erbes. Die Speicherung der Forschungsergebnisse in einem *Repositorium* soll gewährleisten, dass sie nachhaltig verfügbar, verifizierbar, zitierbar und nachnutzbar sind.

FAIR-Prinzipien

Die *FAIR-Prinzipien* formulieren internationale Grundsätze für nachhaltig nachnutzbare *Forschungsdaten*. Hauptziel ist eine optimale Aufbereitung der *Forschungsdaten*, damit diese findbar (**F**indable), zugänglich (**A**ccessible), austauschbar (**I**nteroperable) und nachnutzbar (**R**eusable) sind. Die *FAIR-Prinzipien* wurden von einer breit aufgestellten Interessengruppe, bestehend aus Vertretern aus Wissenschaft, Industrie, Förderorganisationen und wissenschaftlichen Verlagen, erstellt und 2016 erstmals veröffentlicht.

Forschende

Forschende sind alle forschungsaktiven Mitglieder der Technischen Hochschule Köln einschließlich der Promovierenden. Personen, die nicht direkt mit der TH Köln verbunden sind, aber zu Forschungszwecken die Infrastruktur nutzen oder an der TH Köln physisch präsent sind, werden ebenfalls in den Begriff einbezogen. Von Gastforschenden oder externen Kooperationspartnern wird ebenso erwartet, dass sie den Leitlinien entsprechend handeln.

Forschungsdaten

Forschungsdaten sind jegliche Art von Daten, die während des Forschungsprozesses generiert, erhoben oder genutzt werden oder dessen Ergebnis sind, unabhängig von der Art der Entstehung oder des Speichermediums. Dazu zählen u.a. Messdaten, Laborwerte, audiovisuelle Informationen, Texte, Abbildungen, Codes, Objekte aus Sammlungen oder Proben, die in der wissenschaftlichen Arbeit entstehen, entwickelt, erhoben oder ausgewertet werden. Als *Forschungsdaten* werden sowohl Primär-, als auch Sekundärdaten zum jeweiligen Forschungsprojekt angesehen. Ihre Vielfalt spiegelt somit die Vielfalt der wissenschaftlichen Disziplinen, der Erkenntnisinteressen und der Forschungsmethoden wider.

Forschungsdatenmanagement

Unter Forschungsdatenmanagement (FDM) wird der Prozess der Transformation, Selektion und Speicherung von *Forschungsdaten* mit dem Ziel, diese langfristig und unabhängig vom Datenerzeuger zugänglich, nachnutzbar und nachprüfbar zu halten, verstanden. Es können dazu an allen Punkten des Datenlebenszyklus strukturierte Maßnahmen ergriffen werden, um die wissenschaftliche Aussagekraft von *Forschungsdaten* zu erhalten, deren Zugänglichkeit durch Dritte für Auswertung und Analyse zu bewahren und die Nachweiskette zu sichern.

Langzeitarchivierung

Die *Langzeitarchivierung* (LZA) ist ein Verfahren, das Forschungsdaten für einen unbestimmten Zeitraum (i.d.R. > 10 Jahre ab Projektende), über technologische und soziokulturelle Veränderungen hinaus, verfügbar und interpretierbar hält. Dabei gilt es, einem Datenverlust vorzubeugen, die authentische Datenwiedergabe langfristig zu ermöglichen und ein geeignetes Archivierungssystem (*Metadaten*, Struktur) zu verwenden. Dies kann erreicht werden, indem Dateien regelmäßig in aktuelle Dateiformate konvertiert werden, oder indem von vornherein Dateiformate verwendet werden, die offen und gut dokumentiert sind.

Metadaten

Metadaten sind Daten über Daten. Es sind an sich unabhängige Daten, die beschreibende oder kontextuelle Informationen zu anderen Daten enthalten, indiziert werden können und das Archivieren und Auffinden erleichtern. Es gibt unterschiedliche Arten bzw. Kategorien von *Metadaten*. Meist wird zwischen fachlichen und technischen *Metadaten* unterschieden. Während letztgenannte einen klaren *Metadaten*-status haben, können fachlich *Metadaten* bisweilen auch als *Forschungsdaten* begriffen werden.

Persistente Identifikatoren

Ein *Persistent Identifier* (Persistent Identifier; PI) ist eine konstante Internet-Adresse für digitale Objekte. Er stellt sicher, dass ein Datensatz dauerhaft auffindbar, abrufbar und zitierbar bleibt, auch wenn sich sein physischer Standort ändert. Bekannte Beispiele sind DOI (Digital Object Identifier), ORCID (Open Researcher Contributor Identification) oder URN (Uniform Resource Name). Im Gegensatz zu anderen seriellen Identifikatoren (bspw. URL-Adressen) verweist ein Persistent Identifier auf das Objekt selbst und nicht auf seinen Standort im Internet.

Primärdaten

Primärdaten sind i.R. erhobene Rohdaten, die weder bearbeitet, geprüft noch kommentiert bzw. mit *Metadaten* versehen wurden, die aber die Grundlage für die wissenschaftliche Beschäftigung mit einem Gegenstand darstellen. Im Gegensatz dazu stehen die abgeleiteten *Sekundärdaten* (prozessierte Daten), die bei der Rohdatenverarbeitung aus den *Primärdaten* gewonnen werden.

Sekundärdaten

Sekundärdaten sind diejenigen Daten, die aus *Primärdaten* durch Verarbeitungsschritte hervorgehen. Es handelt sich um abgeleitete oder prozessierte Daten, die im Rahmen der *Primärdaten*-verarbeitung als Ergebnis z.B. als Publikation anfallen.