

**Leitfaden zur
„FAIRifizierung“ des
Datenmanagements
und der Ermöglichung
der Nachnutzung
von Daten**

PARTHENOS



FINDABLE



ACCESSIBLE



INTEROPERABLE



REUSABLE

Über diesen Leitfaden

Dieser Leitfaden enthält eine Reihe von Leitsätzen für ein aufeinander abgestimmtes Vorgehen von Datenproduzierenden, Daten-Archivarinnen und -Archivaren und Datennutzenden in den Geistes- und Sozialwissenschaften, um eine größtmögliche Nachnutzbarkeit von Forschungsdaten zu erreichen.

Der Leitfaden basiert auf der Arbeit von über fünfzig Mitgliedern des PARTHENOS Projektes. Ihre Aufgabe bestand darin, Gemeinsamkeiten bei der Implementierung von Leitlinien und Strategien für das Forschungsdatenmanagement zu erforschen und sie benutzen dafür Ergebnisse aus Literaturrecherchen, Fragebögen und Interviews mit ausgewählten Experten um ca. 100 Leitlinien zu sammeln (inklusive expliziter und impliziter Empfehlungen für Formate, Richtlinien zur Datenprüfung und „Best Practices“).

Mit dem Fokus auf die Qualität der (Meta-) Daten und Repositorien erarbeitete das PARTHENOS Team zwanzig Leitsätze, die unterschiedlichen Fachdisziplinen gemeinsam haben.

Zur besseren Orientierung hat das Team jedem der Leitsätze eines der Prinzipien Auffindbarkeit (Findable), Zugänglichkeit (Accessible), Interoperabilität (Interoperable) und Nachnutzbarkeit (Reusable) zugeordnet. Diese Einteilung basiert auf den FAIR Data Prinzipien – 2016 erstmals von FORCE11 veröffentlicht – und soll denjenigen eine Orientierungshilfe bieten, die die Nachnutzung von Forschungsdaten verbessern wollen. In jedem der PARTHENOS Leitsätze finden sich spezifische Empfehlungen einerseits für Datenproduzierende und Datennutzende und andererseits für Datenarchivarinnen und -archivare. Die Symbole unter den Leitsätzen verdeutlichen die jeweils adressierte Zielgruppe.



Das Symbol „Glühbirne“ verweist auf Empfehlungen für Datenproduzierende und Datennutzende, wie zum Beispiel Forschende und Forschungscommunities in den Geschichtswissenschaften, Archäologie, Sprachwissenschaften und Sozialwissenschaften.



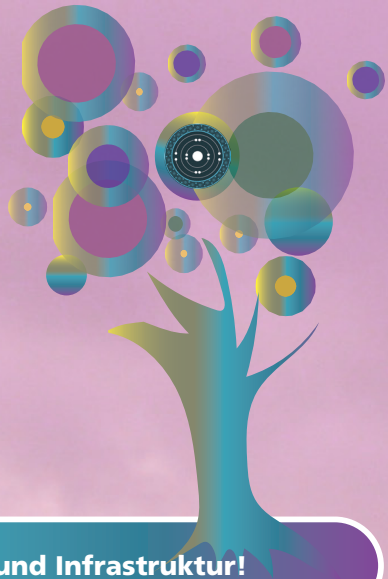
Das Symbol „Zahnräder“ weist auf Empfehlungen für Forschungsinfrastrukturen und Datenarchive in Forschungsinstituten und Kulturerbe-Einrichtungen hin.



PARTHENOS ist ein Konsortium aus sechzehn europäischen Forschungseinrichtungen und Forschungsinfrastrukturen. Die PARTHENOS-Mitglieder wollen die Nachnutzbarkeit von Forschungsdaten verbessern, indem sie Brücken zwischen den Datenlebenszyklen von einzelnen Forschungscommunities, Datenrepositorien, Forschungsinfrastrukturen und Kulturerbe-Einrichtungen in den miteinander eng verzahnten Bereichen der Geistes- und Sozialwissenschaften bauen.

20 LEITLINIEN

für die „FAIRifizierung“ des Datenmanagements und der Ermöglichung der Nachnutzung von Daten



1

Investieren Sie in Menschen und Infrastruktur!

Eine wichtige Voraussetzung für die Umsetzung der weiteren 19 Leitsätze dieses Leitfadens ist die Investition in die Dateninfrastruktur und in die Einstellung und Ausbildung von Datenexpertinnen und -experten.



Machen Sie sich mit Best-Practice-Beispielen im Forschungsdatenmanagement vertraut. Schauen Sie sich die PARTHENOS Trainingsmodule zum Datenmanagement oder den CESSDA Data Management Expert Guide an.



Investieren Sie in die Einstellung und Ausbildung von Datenexpertinnen und -experten und sehen Sie ein Budget für Investitionen in die technische Infrastruktur und Personal vor.

FINDABLE

Forschungsdaten sollten sowohl für Menschen als auch für Computersysteme leicht aufzufinden sein und daher auf verbindlichen Beschreibungen von Metadaten basieren, die das Auffinden relevanter Datensätze ermöglichen.

2

Verwenden Sie Persistent Identifier!

Die Auffindbarkeit von Daten ist eine notwendige Voraussetzung für jeden weiteren Schritt, angefangen von der Zugänglichkeit bis hin zur Nachnutzung. Um auffindbar zu sein, müssen alle Daten und Datensätze eindeutig und über einen Persistent Identifier (PID) langfristig identifizierbar sein. Ein PID funktioniert weiterhin, auch wenn sich die Web-Adresse einer Ressource verändern sollte. PIDs gibt es in zahlreichen Varianten, wie z. B. Handle, DOI, PURL oder URN.



Verwenden Sie in Ihren Forschungsergebnissen PIDs als Verweis auf Ihre Datensätze.



Wählen Sie ein geeignetes beständiges Identifikationsschema aus und weisen Sie jeder Ressource einen PID zu. Verwenden Sie den PID Guide von NCDD, um sich für einen geeigneten PID für Ihre Forschungsinfrastruktur zu entscheiden.

3

Forschungsdaten immer mit Quellenangabe!

Wenn Forschungsdaten einen Persistent Identifier haben und gemäß den fachspezifischen Standards zitiert werden, sind sie leichter aufzufinden.



Machen Sie sich mit den für Ihr Forschungsgebiet bzw. Ihre Fachdisziplin üblichen Zitierweisen von Daten vertraut und zitieren Sie Daten dementsprechend.



Stellen Sie den Forschungscommunities Informationen zu Best-Practice-Beispielen für das Zitieren von Daten zur Verfügung und machen Sie es Datennutzenden leicht, Daten zu zitieren, z. B. durch die Bereitstellung eines anklickbaren Buttons, der einen „Zitervorschlag“ enthält.

4

Verwenden Sie persistente Autorenidentifikatoren!

Ein Persistent Identifier für Autorinnen oder Autoren (z. B. VIAF, ISNI oder ORCID) erleichtert es, Datensätze, Forschungsaktivitäten, Publikationen und Forschende miteinander zu verlinken und erleichtert so die wissenschaftliche Anerkennung und Auffindbarkeit.



Heben Sie sich von anderen Forschenden bzw. Forschergruppen ab. Legen Sie sich einen Autorenidentifikator zu, sofern Sie noch keinen besitzen, und verwenden Sie ihn in ihren Datensätzen als Referenz.



Geben sie Autorenidentifikatoren in den Metadaten an.

5

Wählen Sie ein geeignetes Metadatenschema!

Metadaten sind für die Auffindbarkeit von Daten wesentlich. Das gilt insbesondere für die Metadaten, die für das Zitieren und Beschreiben von Daten verwendet werden. Ein Metadatenschema besteht aus einer Liste von standardisierten Elementen, die Informationen über die Ressource erfassen, z. B. einem Titel, einem Identifikator, einem Urheber oder einem Datum. Die Verwendung von bereits vorhandenen Metadatenschemata stellt die Einhaltung internationaler Standards für den Austausch von Daten sicher.



Beschreiben Sie Forschungsdaten so konsistent und umfassend wie möglich, um die Auffindbarkeit von Inhalten sicher zu stellen. Fügen Sie ausreichend Informationen hinzu, die sicherstellen, dass die Daten auch in der Zukunft zugänglich und verständlich sind. Verwenden Sie möglichst ein bereits vorhandenes Metadatenschema, das für Ihren Objekttyp oder Ihren Datensatz geeignet ist.



Machen Sie unmissverständlich deutlich, welches Metadatenschema Sie verwenden und der Fachcommunity empfehlen. Denken Sie über die Verwendung eines Dateneingabeformulars nach, das weitere Metadaten z. B. zur Provenienz der Daten sammelt, um so die Datensätze beim Speichern zusätzlich anzureichern.

ACCESSIBLE

Forschungsdaten sollten einfach zugänglich sein, klar definierte Zugangsbedingungen haben und unter Verwendung von standardisierten Protokollen abrufbar sein.



Wählen Sie ein vertrauenswürdiges Repository aus!

Ein zertifiziertes Repository bietet einen vertrauenswürdigen Speicherort für Datensätze. Die Zertifizierung ist eine Garantie dafür, dass die Daten sicher aufbewahrt werden und langfristig verfügbar, auffindbar und zugänglich bleiben. Beispiele für Zertifikate sind das CoreTrustSeal, das nestor-Siegel und die ISO 16363 Zertifizierung.



Machen Sie Ihre Daten über ein vertrauenswürdiges Repository zugänglich. So können Sie zugleich sicherstellen – wenn Sie die Standards des Repositoriums für empfohlene Dateiformate, Metadatenschemata usw. einhalten –, dass Sie alle Voraussetzungen erfüllen, um Ihre Daten FAIR zu machen.



Machen Sie den Zertifizierungsgrad auf Ihrer Webseite deutlich. Wenn Sie (noch) nicht zertifiziert sind, geben Sie an, wie Sie die Verfügbarkeit, Auffindbarkeit, Zugänglichkeit, und Nutzbarkeit von Daten langfristig sicherstellen werden.

7

Verwenden Sie für den Datenaustausch Standardprotokolle!

Die Verwendung von standardisierten Übergabeprotokollen ermöglicht es Forschungsinfrastrukturen (Meta-)Daten öffentlich verfügbar und harvestbar („abziehbar“) beispielsweise für Suchmaschinen zu machen, was die Zugänglichkeit erheblich verbessert.



Verwenden Sie Standardprotokolle, wie SWORD, OAI-PMH, ResourceSync und SPARQL. Wandeln Sie Metadatenschemata in XML oder RDF um. Führen Sie ein Verzeichnis der Protokoll-Endpoints und der Pfade über die auf die Forschungsdaten zugegriffen werden kann und machen Sie es öffentlich verfügbar.

8

Machen Sie klare Angaben zur Zugänglichkeit!

Informationen zur Zugänglichkeit verdeutlichen, wie Datennutzende auf Datensätze zugreifen dürfen. Bereits beim Ablegen von Daten in einem Datenrepositorium muss klar ersichtlich sein, unter welchen Zugriffsmöglichkeiten ausgewählt werden kann.



Beachten Sie bei der Auswahl von Zugriffsoptionen rechtliche Auflagen, fachspezifische Grundsätze und ethische Aspekte, sofern diese anzuwenden sind. Entscheiden Sie sich für Open Access, wo immer dies möglich ist. Wenn Sie personenbezogene Daten sammeln, fragen Sie sich, ob diese Informationen enthalten, die Rückschlüsse auf die Identität der Teilnehmenden erlauben. Fragen Sie sich auch, wozu die Beteiligten eingewilligt haben und welche Maßnahmen zum Schutz Ihrer Daten Sie ergriffen haben. Falls Ihre Daten nicht Open Access veröffentlicht werden dürfen, sollten dies zumindest die Metadaten sein, um die Auffindbarkeit sicherzustellen.



Fördern Sie die Open Access Veröffentlichung von (Meta-)Daten. Machen Sie deutlich, welche Einschränkungsmöglichkeiten es für sensible (Meta-)Daten gibt, die nicht öffentlich zugänglich sein sollen. Versuchen Sie in diesem Fall, die (Meta-)Daten durch ein kontrolliertes und dokumentiertes Zugriffsverfahren verfügbar zu machen.

9

Verwenden Sie Sperrfristen für Daten, wo immer dies erforderlich ist!

Während der Daten-Sperrfrist wird nur die Beschreibung des Datensatzes veröffentlicht. Die eigentlichen Daten sind nicht zugänglich. Nach Ablauf der festgelegten Sperrfrist werden die vollständigen (Meta-)Datensätze verfügbar.



Machen Sie deutlich, warum und für welchen Zeitraum die Daten-Sperrfrist erforderlich ist. Stellen Sie die (Meta-)Daten so bald wie möglich zur Verfügung.



Geben Sie an, ob und unter welchen Bedingungen die Einrichtung von Sperrfristen erlaubt ist.

Um das Auffinden zu beschleunigen und neue Einblicke zu ermöglichen, sollten Forschungsdaten von Menschen und auch von Computersystemen möglichst einfach mit anderen Datensätzen verknüpft werden können.

INTEROPERABLE

10

Richten Sie gut dokumentierte und maschinenlesbare APIs ein!

Gut dokumentierte und maschinenlesbare APIs – bestehend aus einer Reihe von Definitionen von Unterprogrammen, Protokollen und Entwicklungswerkzeugen für Anwendungsprogramme – bieten die Möglichkeit zur automatischen Indexierung, zum Abruf und zur Verknüpfung von (Meta-)Daten aus unterschiedlichen Datenrepositorien.



Dokumentieren sie APIs gut und stellen Sie das (Meta-)Daten Modell in übertragbarer Form bereit. Denken Sie darüber nach, Beispiele für eine erfolgreiche Datenanalyse auf der Basis von verschiedenen Endpoints anzugeben und zeigen Sie wie diese in neuen Datensätzen zusammengeführt für neue Forschungsansätze verwendet werden können.

11

Verwenden Sie offene und klar definierten Vokabulare!

Die Beschreibung von Metadatenelementen sollte den Anforderungen der Community entsprechen und offene, klar definierte und bekannte Vokabulare verwenden. Solche Vokabulare beschreiben die genaue Bedeutung der Konzepte und Eigenschaften, die die Daten repräsentieren.



Verwenden Sie die für Ihr Forschungsgebiet relevanten Vokabulare zur Anreicherung und gliedern Sie Ihre Forschungsergebnisse von Beginn Ihres Forschungsprojektes an entsprechend.



Nennen sie Beispiele für Vokabulare, die die Forschungscommunity verwendet und die auf den Besonderheiten des Forschungsbereichs basieren.

12

Dokumentieren Sie Metadatenmodelle!

Die klare Dokumentation von Metadatenmodellen hilft den Entwicklern beim Vergleichen und Mappen von Metadaten.



Veröffentlichen Sie die in ihrer Einrichtung verwendeten Metadatenmodelle. Dokumentieren Sie technische Anforderungen und definieren Sie „Klassen“ (Gruppen von Dingen, die gemeinsame Eigenschaften haben) und „Properties“ (Elemente, die sowohl die Eigenschaften von einzelnen Metadatengruppen als auch die Beziehungen zwischen unterschiedlichen Metadatengruppen beschreiben). Um das Mappen von Metadaten zu vereinfachen, listen Sie verbindliche und empfohlene „Properties“ auf.

13

Schreiben Sie Datenstandards vor, die interoperabel sind und verwenden Sie nur diese!

Die Verwendung eines Datenstandards, der von einer starken Community unterstützt wird, erhöht die Möglichkeit Datenbestände zu teilen, nachzutzen und zusammenzuführen.



Erkundigen Sie sich bei den Betreibern des Repositoriums, dem Sie Ihre Daten übergeben wollen, welche Standards für Daten dort verwendet werden. Strukturieren Sie Ihre Datenerfassung von Projektbeginn an entsprechend.



Schreiben sie vor, welche Datenstandards Ihre Einrichtung verwenden soll, bauen Sie eine entsprechende Community auf und pflegen Sie sie besonders im Hinblick auf Interoperabilität. Gute Beispiele sind CMDI (Sprachwissenschaften) und der SIKB0102 Standard (Archäologie).

14

Etablieren Sie Prozesse zur Verbesserung der Datenqualität!

Um die (Meta-)Datenqualität und dementsprechend die Interoperabilität zu verbessern, sollten (automatisierte) Prozesse zur Bereinigung, Ableitung und Anreicherung von (Meta-)Daten etabliert werden.



Legen Sie Verfahren fest, die das Fehlerrisiko bei der Datenerfassung minimieren. Zum Beispiel können Sie einen Termin im Kalender zur Wahl stellen, um eine händische Eingabe zu vermeiden.



Investieren Sie in Werkzeuge, die Ihnen helfen, (Meta-)Daten zu bereinigen und in standardisierte und interoperable Formate zu konvertieren. Unternehmen Sie gemeinsame Anstrengungen zur Entwicklung von Workflows und Softwarelösungen für diese automatisierten Prozesse, z. B. indem Sie Machine Learning Tools verwenden.

15

Schreiben Sie zukunftssichere Dateiformate vor und verwenden Sie nur diese!

Alle Datensätze in einem Repositorium sollten in einem offenen, international standardisierten Dateiformat vorliegen, um eine langfristige Interoperabilität im Hinblick auf Nutzbarkeit, Zugänglichkeit und Nachhaltigkeit sicher zu stellen.



Denken sie an zukunftsfähige Dateiformate von Beginn Ihres Projektes an. Verwenden Sie gebräuchliche Formate, die von Repositorien empfohlen werden und unabhängig von spezifischer Software, Entwicklern oder Anbietern sind.



Fördern Sie die Verwendung von Formaten, die für die digitale Langzeitarchivierung geeignet sind, wie PDF-A-, CSV- oder MID/MIF-Dateien. Stellen Sie eine einfache zu findende und detaillierte Übersicht über zulässige Dateiformate zur Verfügung.

Forschungsdaten sollten für zukünftige Forschung und weitere Verarbeitung bereitstehen, so dass Ergebnisse selbstverständlich reproduzierbar sind und neue Untersuchungen erfolgreich auf bereits früher gewonnenen Ergebnissen aufbauen können.

REUSABLE

16

Dokumentieren Sie Daten systematisch!

Um deutlich zu machen, was in einem Datensatz oder Repositorium zu erwarten ist und was nicht, sollten Daten systematisch dokumentiert werden. Transparenz darüber, was die Daten beinhalten und was nicht, schafft Vertrauen und führt so letztendlich zur Nachnutzung der Daten.



Stellen Sie Codebücher einschließlich einer Beschreibung der Methodik, einer Liste von Abkürzungen, einer Beschreibung von Lücken in den Daten, der Konfiguration der Datenbank etc. zur Verfügung.

17

Beachten Sie Namenskonventionen!

Das Beachten einer präzisen und konsistenten Namenskonvention – einem allgemein anerkanntes Schema für die Benennung von Datensätzen – erleichtert zukünftigen Generationen von Forschenden den Abruf, Zugriff und das Verständnis von Datenobjekten und Datensätzen erheblich.



Ziehen Sie Leitlinien und bewährte Verfahren Ihrer Fachdisziplin oder Ihres Forschungsfeldes zu Rate, um die am besten geeignete Namenskonvention zu finden.



Machen Sie anhand von in der Praxis bewährten Beispielen deutlich, wie spezifische Namenskonventionen für Daten zu gestalten und anzuwenden sind.

18

Nutzen Sie gängige Dateiformate!

Indem sie standardisierte Dateiformate verwenden, die in Ihrer Community weit verbreitet sind, erhöhen Sie die Nachnutzbarkeit.



Verwenden Sie aktuelle und gängige Dateiformate parallel zu Archivierungsformaten um Ihre Daten zu teilen, z. B. Excel (xlsx) und CSV oder ESRI Shapefiles neben MID/MIF-Dateien.



Veröffentlichen Sie Daten nicht nur in gängigen Formaten, sondern auch in Archivierungsformaten, falls diese unterschiedlich sein sollten.

19

Erhalten Sie die Datenintegrität aufrecht!

Die gesammelten Forschungsdaten sollten mit den Forschungsdaten, auf die später zugegriffen wird, identisch sein. Um die Authentizität der Daten sicher zu stellen, sollten Prüfungen der Datenintegrität durchgeführt werden.



Richten Sie ein Verfahren für die Versionskontrolle ein. Es ist von wesentlicher Bedeutung, dass Sie sicherstellen, dass jede Änderung in einer überarbeiteten Version eines Datensatzes korrekt dokumentiert wird.



Um festzustellen, ob eine Datei verändert wurde, ist es unerlässlich, die Provenienz – d. h. den Ursprung der Daten und jede vorgenommene Veränderung an ihnen – zu erfassen und jede Kopie mit dem Original zu vergleichen. Eine Prüfung der Datenintegrität kann mittels eines Hashwertes (Fingerprint), wie beispielsweise einer Prüfsumme, oder durch den direkten Vergleich zweier Dateien durchgeführt werden. Stellen Sie einen Mechanismus zur Verfügung, um verschiedene Versionen ansprechen zu können, z. B. indem Sie die Versionsnummer als Suchparameter zum Identifier hinzufügen.

20

Vergeben Sie Nachnutzungslizenzen!

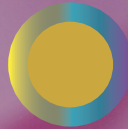
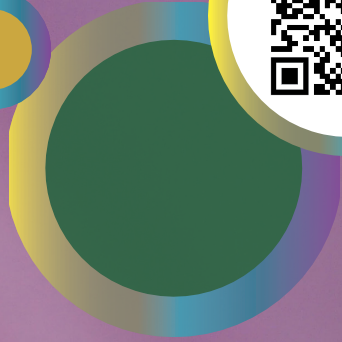
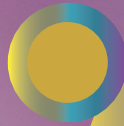
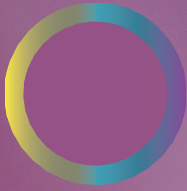
Um die bestmögliche Nachnutzung von (Meta-)Daten zu ermöglichen, sollte klar sein, wer die Rechte an den (Meta-)Daten besitzt und welche Lizenz gilt.



Vergewissern Sie sich, wer die Rechte an den (Meta-)Daten besitzt ehe Sie Ihre Forschungsdaten veröffentlichen.



Machen Sie die (Meta-)Datenlizenz und die Nachnutzungsmöglichkeiten transparent und in einem maschinenlesbaren Format publik. Um die Interoperabilität zu verbessern, versuchen Sie, Ihre Lizenzen auf Rahmenmodelle zu mappen, die bereits erfolgreich eingesetzt werden, wie beispielsweise Creative Commons.



PARTHENOS ist ein durch die Europäische Kommission in der Förderrichtlinie Horizon 2020 gefördertes Projekt. Die in dieser Veröffentlichung geäußerten Ansichten und Meinungen liegen in der alleinigen Verantwortung der Autorinnen und Autoren und geben nicht unbedingt die Meinung der Europäischen Kommission wieder.

Dieser Leitfaden (Version August 2019) steht unter der Creative Commons CC BY 4.0 Lizenz.
Design: Verbeeldingskr8. DOI: 10.5281/zenodo.3363078.