

Richtlinien für die Vorbereitung von Replikationsdateien empirischer Forschung¹

Version 7.0 vom 26. März 2025

Die *Politische Vierteljahresschrift (PVS)*, herausgegeben durch die *Deutsche Vereinigung für Politikwissenschaft (DVPW)*, verpflichtet die Autor:innen aller akzeptierten Manuskripte, ihre Replikationsdaten in einem etablierten sozialwissenschaftlichen Datenrepository bereitzustellen, bevor der Artikel in die Produktionsphase des Publikationsprozesses eintritt. Als einfachsten Weg empfiehlt die Redaktion der PVS, die Replikationsdaten über den GESIS Datenservice „Archivierung Basis“ ([Link](#)) verfügbar zu machen. Eine Handreichung zur Einreichung der Daten bei GESIS findet sich hier ([Link](#)).

Das vorliegende Dokument enthält Richtlinien für die Vorbereitung von Replikationsdateien, darunter sowohl allgemeine Grundsätze als auch spezifische Anforderungen.² Ziel ist es, einen Standard für die Informationen festzulegen, die zur Verfügung gestellt werden müssen, damit transparent wird, wie die Forschungsarbeit, die in der *Politischen Vierteljahresschrift* publiziert ist, repliziert und weiter ausgewertet werden kann. Dieser Standard soll die aktive Nutzung der Replikationsdateien durch interessierte Mitglieder der sozialwissenschaftlichen Forschungsgemeinschaft erleichtern und fördern.

Diese Richtlinien sollen sicherstellen, dass empirische Forschung bezogen auf die genutzten Daten so transparent wie möglich dokumentiert ist. Dabei erkennt die *Politische Vierteljahresschrift* die Heterogenität sozialwissenschaftlicher Arbeitsweisen an. Dementsprechend werden im Folgenden mehrere Sets von Richtlinien angeboten, die jeweils unterschiedliche Anweisungen für verschiedene Arten von Untersuchungen geben. Wissenschaftler:innen, die mehrere Methoden anwenden, müssen die Richtlinien für jeden Methodentyp befolgen, den sie verwenden. Der PVS ist bewusst, dass diese Richtlinien eventuell nicht alle Formen der Datenerhebung und Datenanalyse abdecken, die in der Forschung genutzt werden, oder dass diese Richtlinien aus bestimmten Gründen nicht angewendet werden können. Ausnahmen von den Regeln können erforderlich werden, um den besonderen Umständen einzelner Forschungsarbeiten und datenschutzrechtlichen Bestimmungen Rechnung zu tragen. Der konkrete Umgang ist dann im Einzelfall in Rücksprache zwischen den Autor:innen und der Redaktion zu klären. Unbenommen von dieser Einschränkung bleibt der Anspruch, dass für die große Mehrheit der in der PVS veröffentlichten empirischen Texte zugrundeliegende Daten öffentlich frei verfügbar und replizierbar sind.

¹ Diese Richtlinien sind angelehnt an die „Guidelines for Preparing Replications Files“ des *American Journal of Political Science (AJPS)*, Version 2.1 vom 19. Mai 2016, formuliert von William G. Jacoby und Robert N. Lupton.

² Mitarbeiter:innen, angeschlossene Organisationen und Benutzer:innen des Qualitative Data Repository (Syracuse University) lieferten die Richtlinien für qualitative Analysen und leisteten eine Reihe weiterer wichtiger Beiträge für dieses Dokument. Wir sind sehr dankbar für ihre engagierte Arbeit, ihre redaktionellen Fähigkeiten und ihre starke Unterstützung.

1 ALLGEMEINE GRUNDSÄTZE

Alle Analysen, über die in PVS-Artikeln berichtet wird, sollen öffentlich verfügbar und für die wissenschaftliche Forschungsgemeinschaft zugänglich sein. Autor:innen sollen alle Informationen und Daten zur Verfügung stellen, die für die Reproduktion der publizierten Analyseergebnisse bzw. der publizierten schlussfolgernden/interpretierenden Aussage erforderlich sind. Autor:innen sollen angeben, dass eine informierte Einwilligung der Beteiligten bei Befragungen, Experimenten und Interviews vorliegt. Darüber hinaus sollen Autor:innen angeben, welche Zugangskategorie sie zu ihren Daten gewähren (bspw. freuer Zugang mit oder ohne Registrierung beim Datenrepository, eingeschränkter Zugang oder eine Embargoerfrist). Gründe für Ausnahmen von der Öffentlichkeitsforderung sind im Abschnitt „4. Mögliche Ausnahmen von den Datenzugriffsanforderungen“ beschrieben.

Der Titel der Dateneinreichung muss lauten: „Data: [Titel des Manuskripts]“. Eine kurze Abstract-Beschreibung der eingereichten Daten im Datenrepository ist gewünscht.

2 ANWEISUNGEN UND EMPFEHLUNGEN FÜR DIE DOKUMENTATION VON QUANTITATIVEN ANALYSEN

In den meisten Fällen werden die Replikationsdateien Folgendes enthalten:

- eine Readme-Datei,
- den Analysedatensatz,
- den Code für die Ausführung der betreffenden Software,
- ein Codebuch.

Beachten Sie bitte, dass die Replikationsmaterialien alle im Hauptartikel und in den Rahmeninformationen des Artikels (Anhänge) vorgestellten Analysen abdecken müssen. Den Autor:innen steht es aber frei, Erweiterungen zu denen im Artikel beschriebenen Analysen hinzuzufügen. Auf jeden dieser Punkte wird im Folgenden ausführlicher eingegangen werden.

2.1 Readme-Datei

Jeder Eintrag in einem etablierten Datenrepository muss eine Klartextdatei mit dem Namen <readme.txt> enthalten. Diese Datei enthält den Titel des replizierten Artikels sowie die Namen aller anderen Dateien, die im Replikationspaket enthalten sind, zusammen mit einer kurzen Beschreibung jeder einzelnen Datei. Bei einer kleinen Zahl der im Datenrepository hinterlegten Dateien können diese Informationen in Form einer einfachen Liste dargestellt werden. Der Titel der Dateneinreichung muss lauten: „Data: [Titel des Manuskripts]“.

2.2 Analysedatensatz/-datensätze

Jeder Eintrag im Datenrepository-Datensatz muss eine oder mehrere Dateien mit den Daten enthalten, die erforderlich sind, um die Analyseergebnisse zu reproduzieren, über die im PVS-Artikel und seinen Anhängen berichtet wird. Die Daten müssen entweder als txt-Dateien oder als csv-Dateien im Datenrepository abgelegt werden, um für Forscher:innen in der sozialwissenschaftlichen Community hinreichend leicht zugänglich zu sein.

Die Daten sollen innerhalb jeder Datei so angeordnet werden, dass das Verständnis des Inhalts erleichtert wird. Wenn möglich, sollen die Variablen aussagekräftige Namen erhalten. Außerdem soll eine eindeutige FallkennzeichnungsvARIABLE enthalten sein. Wenn die Daten aus einer anderen Quelle extrahiert werden, dann soll die Fallkennung im Analysedatensatz diesen Fall mit der Information in der ursprünglichen Quelle verknüpfen.

Bei Analysen, die auf sehr datenintensiven Verfahren basieren (z. B. Bayessche Simulationen, Bootstrap Resampling usw.), ist es nicht erforderlich, den vollständigen Inhalt jedes einzelnen replizierten Datensatzes verfügbar zu machen. Der vollständige Satz relevanter Ergebnisse (z. B. die simulierten Parameterwerte bei einer MCMC-Schätzung, die Bootstrap-Replikationen einer Stichprobenstatistik usw.) soll jedoch in einer kohärenten Datendatei zur Verfügung gestellt werden. In solchen Fällen ist die Bereitstellung von Software-Befehlsdateien (siehe nächster Unterabschnitt) zur Reproduktion der gesamten datenintensiven Analyse besonders wichtig.

In einigen Fällen (siehe Abschnitt „Mögliche Ausnahmen von den Datenzugriffsanforderungen“) können die Analysen in einem *PVS*-Artikel auf Daten beruhen, die nicht an einem öffentlich zugänglichen Ort verfügbar gemacht werden können, weil Rechte Dritter betroffen sind. Auf diesen Umstand muss bei der ersten Einreichung des Manuskripts von den Autor:innen ausdrücklich hingewiesen werden, damit die Redaktion über eine Ausnahme von der Replikationsrichtlinie entscheiden kann. Sobald diese Genehmigung vorliegt, muss der Analysedatensatz nicht mehr bei einem etablierten Datenrepository eingestellt werden. Die Autor:innen müssen jedoch immer noch Anweisungen (do-files, R-scripts etc.) bereitstellen, die von interessierten Forscher:innen für den Zugriff auf die Daten verwendet werden können (siehe Abschnitt „Informationen zur Reproduktion des Analysedatensatzes“ weiter unten), sowie Formatierungs- und Variablendefinitionsinformationen für die Daten, die im *PVS*-Artikel analysiert werden.

2.3 Befehlsdateien (Skripte)

Jeder Datensatz muss eine oder mehrere Dateien (Skripte) mit den Anweisungen enthalten, die auf den Analysedatensatz angewendet werden können, um alle im *PVS*-Artikel vorgestellten Analyseergebnisse zu reproduzieren. In der Regel handelt es sich dabei um reine Textdateien, das genaue Format der Dateiinhalte hängt von der Software ab, die für die Durchführung der ursprünglichen Analysen verwendet wurde.

Unabhängig vom Format, das für die Befehlsdateien verwendet wird, sollen in den Dateien Kommentare verwendet werden, die die einzelnen Schritte der Analyse erläutern. Die Autor:innen können davon ausgehen, dass die Benutzer:innen mit dem für die Analyse verwendeten Softwaresystem vertraut sind (d. h. es ist nicht notwendig zu erklären, wie bestimmte Befehle funktionieren). Die Autor:innen sollen jedoch erklären, wie die verschiedenen Befehle Ergebnisse erzeugen, die für die im Artikel berichteten Analysen relevant sind (z. B. „Die folgenden Befehle rekodieren die Variablen X und Y in Vorbereitung auf das logistische Regressionsmodell“, „Die folgenden Befehle erzeugen Abbildung 1 im Artikel“ usw.).

In einigen Fällen kann die Durchführung der Analysen in einem *PVS*-Artikel Software-Tools erfordern, die Forschenden nicht ohne weiteres zur Verfügung stehen. Beispiele hierfür sind (beschränken sich aber nicht auf) Stata .ado-Dateien oder R-Pakete, die von dem:r Autor:in geschrieben wurden oder speziell für den:die Autor:in verfügbar sind. Alle derartigen Softwareressourcen, die zur Replikation einer Analyse aus dem *PVS*-Artikel erforderlich sind, müssen in den Replikations-Datensatz für den Artikel aufgenommen werden, zusammen mit der entsprechenden Dokumentation und Anweisungen zur Installation (falls erforderlich) und Verwendung.

Autor:innen sollen immer eindeutige Informationen über die Version des Softwaresystems liefern, das für die Durchführung der in ihrem *PVS*-Artikel berichteten Analysen verwendet wurde. Diese Anforderung ist von entscheidender Bedeutung, da sich Algorithmen, Prozeduren und Funktionen über Software-Versionen hinweg ändern können. Zum Beispiel könnte ein R-Skript mit der Kommentaranweisung „Die folgenden Analysen wurden mit R Version 3.2.2 durchgeführt“ beginnen, oder die Readme.txt-Datei für den Datenrepository-Datensatz könnte die folgende Anweisung enthalten: „Alle Datenanalysen in diesem Artikel wurden mit Stata/MP 14.1 für Windows (64-Bit x86-64) durchgeführt“.

Wie bereits erwähnt, sind Skriptdateien besonders wichtig für datenintensive Analysen, bei denen die „intermediären“ Datensätze, die zur Erzielung der Endergebnisse verwendet werden (z. B. MCMC-Simulationen oder Bootstrap-Replikationen der Originaldaten), selbst nicht zu den Replikationsdateien des Datenrepository-Datensatzes gehören. In solchen Fällen würden die Benutzer:innen die ursprünglichen Analysen neu generieren. Daher muss die Befehlsdatei dafür besonders klare Anweisungen enthalten. Falls zum Beispiel bei der Erzeugung von Zufallszahlen Startwerte (seeds) benutzt werden (z. B. für Monte-Carlo-Simulationen, Bootstrap-Resampling, Jittering-Punkte in einer grafischen Darstellung usw.), sind diese mit anzugeben, um konsistente Ergebnisse bei Replikationen zu gewährleisten.

Die Autor:innen können entweder eine einzige Befehlsdatei bereitstellen, die alle Analysen abdeckt, die in dem *PVS*-Artikel abgehandelt werden, oder aber separate Dateien für die einzelnen Analysen, die in dem Artikel enthalten sind. Im ersten Fall sollen innerhalb der Befehlsdatei Kommentaranweisungen verwendet werden, um die Befehle zu unterscheiden, die für die verschiedenen Analyseergebnisse verwendet werden. Dem Prozess ist am besten gedient, wenn die Autor:innen aussagekräftige Dateinamen anstelle von generischen Dateinamen verwenden. Beispielsweise sind Dateinamen wie `<Logistische_Regressionen.do>` oder `<Reproduktion_von_Abbildung_1.R>` besser geeignet als `<Befehle.do>` oder `<R-Skripte.R>`. Bei externen Datensätzen sollten die Originalnamen der Datensätze beibehalten werden.

2.4 Informationen zur Reproduktion des Analysedatensatzes (Codebücher, Erhebungsinstrumente)

Jeder Replikations-Datensatz muss vollständige Informationen zum Aufbau des/der Analyse-Datensatzes/Datensätze aus den ursprünglichen Datenquellen enthalten. Die genauen Materialien dafür hängen von der Art und den Quellen der Analysedaten ab. Unabhängig von den spezifischen Details müssen interessierte Forscher:innen jedoch immer in der Lage sein, nach den Anweisungen des:r Autor:in die genauen Datenwerte zu reproduzieren, die für alle im *PVS*-Artikel berichteten Analysen verwendet wurden.

Das zugehörige Codebuch soll immer Informationen für alle in der Analyse verwendeten Variablen enthalten. Wenn die Daten in einer Textdatei gespeichert sind, müssen Formatierungsinformationen angegeben werden. Bei selbst erhobenen Daten ist darauf zu achten, dass das Erhebungsinstrument (Fragebogen) ebenfalls dokumentiert wird. Bei der Sekundärnutzung von Daten ist zumindest ausdrücklich darauf zu verweisen, wo weitere Informationen zur Erhebung (Fragebogen, Erhebungsanweisungen etc.) gefunden werden können. Werden beispielsweise Variablen aus anderen Quellendatensätzen mit eigenen Daten zusammengespielt, so müssen die Autor:innen die Auswahl, Operationalisierung und das Zusammenspielen dieser Variablen in geeigneter Weise dokumentieren (etwa durch ein Skript). Ebenso sind die genutzte Version des Quelldatensatzes und das Datum des Zugriffs anzugeben, idealerweise unter Angabe eines persistenten Identifikators (z.B. DOI).

2.5 Automatisierte Inhaltsanalysen

Autor:innen, die eine automatisierte Inhaltsanalyse anwenden, müssen das Verfahren der Textsammlung beschreiben. Idealerweise sollten die Textsammlungen ebenfalls hinterlegt werden (wenn dies rechtlich möglich ist) oder es kann auf bereits existierende Textdatenbanken verwiesen werden. In allen Fällen sollen Autor:innen die document term matrix der Textsammlung in das Replikationsmaterial aufnehmen. Darüber hinaus sollen Autor:innen alle Textverarbeitungsschritte (text pre-processing) beschreiben, die vor der Analyse durchgeführt wurden, und einen Code für die jeweils verwendete Textanalysemethode angeben.

Werden zur Erstellung einzelner Variablen Wörterbuchdateien (dictionaries) oder Codebücher verwendet, so müssen diese nachvollziehbar beschrieben werden (ggf. auch im Anhang der Studie), um die Replizierbarkeit für andere Forscher:innen zu gewährleisten. Die Codebücher müssen alle für die Analysen verwendeten Variablennamen und eine Beschreibung des Variableninhalts enthalten.

3 ANWEISUNGEN UND EMPFEHLUNGEN FÜR DIE DOKUMENTATION VON QUALITATIVEN ANALYSEN

Die potenzielle Vielfalt von qualitativen Analysen impliziert, dass die von den Forscher:innen verwendeten Informationen wahrscheinlich vielfältigere Formate umfassen als Daten, die in quantitativen Analysen verwendet werden. Entsprechend unterschiedlich sind die Anforderungen an die Dokumentation. In jedem Fall müssen die Autor:innen ihre Forschungsprozesse so explizit und präzise wie möglich dokumentieren, um zu veranschaulichen, wie sie zu ihren Ergebnissen und Schlussfolgerungen gelangt sind.

3.1 Matrix-Ansatz für qualitative Daten (QCA und Ähnliches)

3.1.1 Readme-Datei

Genau wie bei quantitativen Analysen muss jede Einreichung im Replikations-Repository eine Klartextdatei mit dem Namen <readme.txt> enthalten. Diese Datei enthält die Namen aller anderen Dateien, die im Replikations-Datensatz enthalten sind, zusammen mit einer kurzen Beschreibung jeder einzelnen Datei. Bei relativ kleinen Replikations-Datensätzen können diese Informationen in Form einer einfachen Liste dargestellt werden. Größere Replikations-Datensätze sollen ähnliche Dateien unter Unterüberschriften gruppieren.

3.1.2 Analysedatensatz/-datensätze

Der „Rohdatensatz“, der aus qualitativen oder quantitativen Quellen stammen kann und anhand von Kalibrierungsschwellen kalibriert wurde, muss zu Replikationszwecken im Datenrepository hochgeladen werden. Die Rohdaten müssen zusammen mit den Kalibrierungsschwellen und -verfahren transparent gemacht werden, sodass die Kalibrierung repliziert werden kann. Eine Ausnahme von der Publikation der eigentlichen Rohdaten ist möglich, wenn der Veröffentlichung der Rohdaten gewichtige Gründe entgegenstehen (bspw. Zusicherung der Vertraulichkeit oder große qualitative Datenmengen). Hier müssen aber die interpretativen Verfahren und Kriterien, die zur Zuordnung der Mengenmitgliedschaft geführt haben, transparent gemacht werden. Bei rein qualitativen nicht-numerischen Rohdaten empfiehlt sich zusätzlich die Veröffentlichung des kalibrierten Datensatzes.

Die Autor:innen müssen den Code für die Kalibrierung der Datensätze, die Konstruktion der Wahrheitswerttabellen, die logische Minimierung der Wahrheitswerttabellen und die (grafische) Darstellung der Ergebnisse bereitstellen. Die zur Verfügung gestellten Materialien sollen Informationen umfassen über:

- a) Die zur Analyse verwendete Software/Package. Zu Beginn des Files sollte alle von den Autor:innen verwendeten Ados/Packages, die in der Syntax verwendet werden, aufgelistet werden.
- b) Die Kalibrierungsfunktion und die Kalibrierungsschwellenwerte mit Begründung: Welche empirischen Informationen führen in welchem Datensatz zu welchem Mitgliedschaftswert [membership score]?
- c) Die Wahrheitsabelle für alle untersuchten Outcomes.
- d) Die Konsistenzschwelle: Ab welchem kleinsten Konsistenzwert wird eine Wahrheitswerttabelle als ausreichend für das Ergebnis angesehen? Die Wahl der Konsistenzschwelle und der dafür benutzten Kriterien sollte begründet werden (z.B. PRI-Werte, widersprüchliche Fälle).
- e) Die Häufigkeitsschwelle: Wie groß muss die Zahl an Fällen in einer Zeile einer Wahrheitswerttabelle mindestens sein, damit diese nicht als „logical remainder“ betrachtet wird?
- f) Die Behandlung von logical remainders: Es soll dokumentiert werden, ob und wenn ja mit welchen kontrafaktischen Methoden logical remainders in die Minimierung der Wahrheitswerttabelle aufgenommen wurden. Die kontrafaktischen Argumente sollten transparent gemacht

werden, bspw. mittels einer Beschreibung und Begründung der kontrafaktischen Annahmen und/oder einer Auflistung der logical remainders, welche in die logische Minimierung mit einbezogen wurden.

- g) Die vollständigen Resultate für Notwendigkeit (sofern anwendbar) und Hinlänglichkeit inklusive der Gütemaße (Konsistenz, Abdeckung, und sofern anwendbar RoN / PRI) und relevante typische und abweichende Fälle (wo anwendbar). Insbesondere:
- h) Sofern mehr als ein Typ von hinlänglichen Lösungsformeln kalkuliert wurde, oder falls es Modellambiguität gibt, sollten alle möglichen Modelle im Online-Appendix aufgeführt werden und die Auswahl des zur Interpretation in der Studie verwendeten Modells kurz begründet werden.
- i) Jede andere Erwägung, die die Entscheidung beeinflusst: Dabei kann es sich um spezifische Fallkenntnisse handeln (z. B. über den standardisierten Indikator, der zur Kalibrierung verwendet wird, wenn ein spezifischer Fall falsch gemessen zu sein scheint). Autor:innen, die Fallkenntnisse nutzen, um solche Messfehler zu korrigieren, sollen in den Bericht aufnehmen, dass und bei welchen Fällen sie so verfahren sind.
- j) Die Resultate der Robustheitsüberprüfung, falls anwendbar.

Die obigen Anforderungen basieren auf der Annahme, dass die Autor:innen das statistische Analyseprogramm R zur Durchführung der QCA verwenden. Autor:innen, welche nicht sicher sind, ob sie die erforderlichen Informationen bereitstellen können (bsp. aufgrund der Funktionalität der verwendeten Software), können sich bei Fragen zu den Materialien und Informationen, die bereitgestellt werden müssen, an den:die PVS-Redakteur:in wenden.

3.1.3 Befehlsdateien (Skripte)

Jeder Datenrepository-Datensatz muss eine oder mehrere Dateien (Skripte) mit den Anweisungen enthalten, die auf den Analysedatensatz angewendet werden können, um alle im PVS-Artikel vorgestellten Analyseergebnisse zu reproduzieren. In der Regel handelt es sich dabei um reine Textdateien, das genaue Format der Dateiinhalte hängt von der Software ab, die für die Durchführung der ursprünglichen Analysen verwendet wurde.

Unabhängig vom Format, das für die Befehlsdateien verwendet wird, sollen in den Dateien ausführliche Kommentare verwendet werden, die die einzelnen Schritte der Analyse erläutern. Die Autor:innen können davon ausgehen, dass die Benutzer:innen mit dem für die Analyse verwendeten Softwaresystem vertraut sind (d. h. es ist nicht notwendig zu erklären, wie bestimmte Befehle funktionieren). Die Autor:innen sollen jedoch erklären, wie die verschiedenen Befehle Ergebnisse erzeugen, die für die im Artikel berichteten Analysen relevant sind (z. B. „Die folgenden Befehle rekodieren die Variablen X und Y in Vorbereitung auf die Kalibrierung des Datensatzes“, „Die folgenden Befehle erzeugen Abbildung 1/Tabelle 1 im Artikel“ usw.).

In einigen Fällen kann die Durchführung der Analysen in einem PVS-Artikel Software-Tools erfordern, die Forschenden nicht ohne weiteres zur Verfügung stehen. Beispiele hierfür sind (beschränken sich aber nicht auf) Stata .ado-Dateien oder R-Pakete, die vom Autor: von der Autorin geschrieben wurden oder speziell für den Autor: die Autorin verfügbar sind. Alle derartigen Softwareressourcen, die zur Replikation einer Analyse aus dem PVS-Artikel erforderlich sind, müssen in den Datenrepository-Datensatz für den Artikel aufgenommen werden, zusammen mit der entsprechenden Dokumentation und Anweisungen zur Installation (falls erforderlich) und Verwendung.

Die Autor:innen können entweder eine einzige Befehlsdatei bereitstellen, die alle Analysen abdeckt, die in dem PVS-Artikel abgehandelt werden, oder aber separate Dateien für die einzelnen Analysen, die in dem Artikel enthalten sind. Im ersten Fall sollen innerhalb der Befehlsdatei Kommentaranweisungen verwendet werden, um die Befehle zu unterscheiden, die für die verschiedenen Analyseergebnisse verwendet werden. Dem Prozess ist am besten gedient, wenn die Autor:innen aussagekräftige

Dateinamen anstelle von generischen Dateinamen verwenden. Beispielsweise sind Dateinamen wie <Kalibrierung.do> oder <Reproduktion_von_Abbildung_1/Tabelle_1.R> besser geeignet als <Befehle.do> oder <R-Skripte.R>.

3.1.4 Informationen zur Reproduktion des Analysedatensatzes (Codebücher, Erhebungsinstrumente)

Jeder Datenrepository-Datensatz muss vollständige Informationen zum Aufbau des/der Analyse-Datensatzes/Datensätze aus den ursprünglichen Datenquellen enthalten. Die genauen Materialien dafür hängen von der Art und den Quellen der Analysedaten ab. Unabhängig von den spezifischen Details müssen interessierte Forscher:innen jedoch immer in der Lage sein, nach den Anweisungen des Autors:der Autorin die genauen Datenwerte zu reproduzieren, die für alle im PVS-Artikel berichteten Analysen verwendet wurden.

Jeder Datei, die einen Analysedatensatz enthält, muss eine pdf-Datei mit einem Codebuch für den Datensatz beigefügt werden. Das Codebuch soll immer Variablendefinitionsinformationen für alle in der Analyse verwendeten Variablen enthalten. Wenn die Daten in einer Textdatei gespeichert sind, müssen Formatierungsinformationen angegeben werden. Bei selbst erhobenen Daten ist darauf zu achten, dass das Erhebungsinstrument (Fragebogen) ebenfalls dokumentiert wird. Bei der Sekundärnutzung von Daten ist zumindest ausdrücklich darauf zu verweisen, wo weitere Informationen zur Erhebung (Fragebogen, Erhebungsanweisungen etc.) gefunden werden können. Werden beispielsweise Variablen aus anderen Quellendatensätzen mit eigenen Daten zusammengespielt, so müssen die Autor:innen die Auswahl, Operationalisierung und das Zusammenspielen dieser Variablen in geeigneter Weise dokumentieren (etwa durch ein Skript). Ebenso sind die genutzte Version des Quelldatensatzes und das Datum des Zugriffs anzugeben.

3.2 Standardisierte Inhaltsanalyse und Dokumentenanalyse

3.2.1 Readme-Datei

Genau wie bei quantitativen Analysen muss jeder Eintrag im Datenrepository eine Klartextdatei mit dem Namen <readme.txt> enthalten. Diese Datei enthält die Namen aller anderen Dateien, die im Datenrepository-Datensatz enthalten sind, zusammen mit einer kurzen Beschreibung jeder einzelnen Datei. Bei relativ kleinen Datenrepository-Datensätzen können diese Informationen in Form einer einfachen Liste dargestellt werden. Größere Datenrepository-Datensätze sollen ähnliche Dateien unter Unterüberschriften gruppieren.

3.2.2 Analysedaten

Jeder Eintrag im Datenrepository-Datensatz muss eine oder mehrere Dateien mit den Daten enthalten, die erforderlich sind, um Analyseergebnisse nachvollziehen zu können, die im PVS-Artikel und seinen Anhängen berichtet werden. Die Daten müssen entweder als txt-Dateien, als csv-Dateien, als pdf-Dateien oder als Dateien weit verbreiteter Textverarbeitungsprogramme im Datenrepository abgelegt werden, um für Forscher:innen in der sozialwissenschaftlichen Community hinreichend leicht zugänglich zu sein.

Die Datei soll so gestaltet sein, dass das Verständnis des Inhalts erleichtert wird. Verschiedene Interviews sind zum Beispiel zu nummerieren, und die Zuordnung zu Bezügen im PVS-Artikel müssen eindeutig sein. Vollständige Transkriptionen – etwa bei Interviews oder bei der Auswertung von medialen Texten – sind nicht notwendig, aber zentrale Stellen, auf die Bezug genommen wird, sind zu dokumentieren.

In einigen Fällen (siehe Abschnitt „Mögliche Ausnahmen von den Datenzugriffsanforderungen“) können die Analysen in einem PVS-Artikel auf Daten beruhen, die nicht an einem öffentlich zugänglichen

Ort verfügbar gemacht werden können, weil Rechte Dritter betroffen sind. Auf diesen Umstand muss bei der ersten Einreichung des Manuskripts von den Autor:innen ausdrücklich hingewiesen werden, damit die Redaktion über eine Ausnahme von der Replikationsrichtlinie entscheiden kann. Sobald diese Genehmigung vorliegt, muss der Analysedatensatz nicht mehr in einem etablierten Datenrepository eingestellt werden. Die Autor:innen müssen jedoch immer noch Anweisungen bereitstellen, die von interessierten Forscher:innen für den Zugriff auf die Daten verwendet werden können (siehe Abschnitt „Informationen zur Reproduktion des Analysedatensatzes“ weiter unten), sowie Formatiereungs- und Variablendefinitionsinformationen für die Daten, die im *PVS*-Artikel analysiert werden.

3.2.3 Informationen zur Reproduktion des Analysedatensatzes (Codebücher, Leitfäden, Fragebögen)

Jeder Datenrepository-Datensatz muss vollständige Informationen zum Aufbau des/der Analyse-Datensatzes/Datensätze aus den ursprünglichen Datenquellen enthalten. Die genauen Materialien dafür hängen von der Art und den Quellen der Analysedaten ab. Unabhängig von den spezifischen Details müssen interessierte Forscher:innen jedoch immer in der Lage sein, nach den Anweisungen des Autors:der Autorin die genauen Datenwerte zu reproduzieren, die für alle im *PVS*-Artikel berichteten Analysen verwendet wurden.

Jeder Datei, die einen Analysedatensatz enthält, muss eine pdf-Datei mit einem Codebuch für den Datensatz beigefügt werden. Das Codebuch soll immer Variablendefinitionsinformationen für alle in der Analyse verwendeten Variablen enthalten. Wenn die Daten in einer Textdatei gespeichert sind, müssen Formatierungsinformationen angegeben werden. Bei selbst erhobenen Daten ist darauf zu achten, dass das Erhebungsinstrument (Fragebogen, Leitfaden) ebenfalls dokumentiert wird. Autor:innen sind darüber hinaus angehalten, die Kontaktaufnahme mit Befragten zu beschreiben und Interviewtranskripte (zumindest der für den Artikel wesentlichen Stellen) bereitzustellen.

Bei Dokumentenanalysen müssen die Autor:innen die Kriterien zur Auswahl der Dokumente detailliert und nachvollziehbar beschreiben. Beispielsweise müssen bei einer Schlagwortsuche, die der Eingrenzung des Datenmaterials diente, diese Schlagworte präsentiert und diskutiert werden.

Autor:innen, die Textdaten mit computergestützter Software zur qualitativen Datenanalyse (CAQDAS) analysieren, müssen eine Liste von Codierungskategorien/-kriterien, eine Beschreibung des Codierungsverfahrens, finale Codierungsbäume und Berichte/Ausgaben bereitstellen.

4 AUSNAHMEN VON DEN DATENZUGRIFFSANSFORDERUNGEN

Das Ziel der Offenlegung einiger oder aller Daten, die der Analyse in einem *PVS*-Artikel zugrunde liegen, kann sich im Widerspruch zu anderen Zielen befinden. Unter diesen Umständen verlangt die Redaktion der *Politischen Vierteljahresschrift* von den Autor:innen, dass sie explizit darlegen, wie sie diese konkurrierenden Anliegen gegeneinander abwägen und wie sich diese Entscheidung auf die Verfügbarkeit der Daten auswirkt, aus denen die Schlussfolgerungen in dem Artikel gezogen werden. Diese Erklärung muss als Teil des Datensatzes des Artikels in das Datenrepository aufgenommen werden.

Die folgenden zwei Arten von Situationen könnten eine Ausnahme rechtfertigen:

4.1 Datensätze mit beschränktem Zugriff

Die erste Situation, für die eine Ausnahme von den allgemeinen Replikationsanforderungen gewährt werden kann, tritt ein, wenn einige oder alle Analysedatensätze (oder zitierten Daten) unter rechtlichen Beschränkungen stehen. Beispiele hierfür sind die Verwendung proprietärer Datensätze, vertragliche Vereinbarungen, die die gemeinsame Nutzung von Daten einschränken, oder Informationen, die aufgrund von Regierungsvorschriften als geheim eingestuft sind. Selbst in solchen Situationen werden

die Autor:innen ermutigt, mit den Inhaber:innen des eingeschränkten Quelldatensatzes über den allgemeinen Zugang zu dem Analysedatensatz zu verhandeln. Wenn der Zugriff nicht gestattet ist, muss der:die Autor:in als Teil des Antrags an die PVS-Redaktion eine Ausnahme von den allgemeinen Replikationsanforderungen beantragen, die folgendes beinhaltet:

- a) Eine Aufstellung der Bedingungen, unter denen der Zugang für qualifizierte Forscher:innen gewährt wird.
- b) Eine Erläuterung der Bedingungen, unter denen die Daten zugänglich gemacht werden.
- c) Die Voraussetzungen und alle anwendbaren Gebühren, die ein Forscher:eine Forscherin vorweisen muss, um Zugang zu den Daten zu erhalten.

4.2 Schutz der beteiligten Personen

Der öffentliche Zugang zu einer Analyse oder einem Quelldatensatz kann aufgrund von Bedenken hinsichtlich des Schutzes von Personen eingeschränkt sein. Den Herausgeber:innen der *Politischen Vierteljahrsschrift* ist bewusst, dass Replikationserfordernisse und Forschungstransparenz im Allgemeinen an ethische Auflagen gebunden sind. An erster Stelle steht dabei die Forderung, dass die Sicherheit, Würde und das allgemeine Wohlbefinden aller Proband:innen gewahrt und geschützt werden müssen. Eine Ausnahme von den allgemeinen Replikationserfordernissen kann beantragt werden, wenn die Freigabe der Analyse oder der Quelldaten diesem Ziel abträglich wäre.

In einigen Fällen können Daten, deren Zugang andernfalls eingeschränkt wäre, durch die Anwendung von Mechanismen, die die Interessen der Proband:innen schützen, zur Verfügung gestellt werden. Zum Beispiel können Autor:innen die Beobachtungen in ihren Daten de-identifizieren und dadurch anonymisieren. Oder die Autor:innen können beschränken, wer die Daten einsehen kann oder wo die Daten eingesehen werden können (z. B. in einer geschützten physischen oder digitalen Datenenklave). Die Mechanismen, die für solche Zwecke zur Verfügung stehen, hängen von der Art der Daten und von den Umständen ab, unter denen die Daten erzeugt wurden. In einigen Situationen stehen solche Mechanismen möglicherweise nicht zur Verfügung. Wenn sie jedoch eine praktikable Option darstellen, werden die Autor:innen ermutigt, sie wo immer möglich einzusetzen.

Wenn die PVS-Redaktion die Genehmigung erteilt, einen Teil oder die Gesamtheit des Analysedatensatzes, des Quelldatensatzes oder der zitierten Daten zurückzuhalten, müssen die freigestellten Informationen nicht in das Datenrepository hochgeladen werden. Der Autor:die Autorin muss zu Beginn des veröffentlichten Artikels einen Hinweis aufnehmen, in dem er:sie ausdrücklich die Einschränkungen der Datenverfügbarkeit hervorhebt, die den öffentlichen Zugang zu den freigestellten Daten verhindern. Unabhängig davon müssen im Datenrepository Materialien bereitgestellt werden, die die Verfahren spezifizieren, mit denen die Daten erhoben und analysiert wurden. Bei Artikeln, die Daten im Matrixformat (quantitativ oder qualitativ) analysieren, muss der Autor:die Autorin relevante Informationen über Variablendefinitionen und Formatierung sowie Analyseskripte für die Durchführung der Analysen und die Konstruktion des Analysedatensatzes bereitstellen. Generell bleiben die Autor:innen verpflichtet, alle anderen Dimensionen der Forschungstransparenz zu befolgen, so dass die Leser:innen umfassend über die Datenanalyse informiert werden, auch wenn die Daten selbst nicht unmittelbar über den Datensatz auf dem Datenrepository verfügbar sind.

5 SCHLUßBEMERKUNGEN

Die vorstehenden Richtlinien beschreiben die Mindestanforderungen für einen Datensatz in einem etablierten Datenrepository. Autor:innen werden aber nicht darauf beschränkt, nur die in diesem Dokument beschriebenen Dateien und Informationen zur Verfügung zu stellen. Vielmehr werden die Autor:innen ermutigt, so viele Informationen wie möglich zur Verfügung zu stellen. Zusätzliche Inhalte des Datensatzes in einem etablierten Datenrepository könnten ergänzende Berichte, Voranalysepläne,

zusätzliche Daten und Erweiterungen der Analyse umfassen, die über die im PVS-Artikel und den veröffentlichten unterstützenden Informationen hinausgehen. Autor:innen werden ermutigt, zu bedenken, dass wissenschaftliche Forschung ein fortlaufender Prozess ist. Es bleibt zu hoffen, dass die auf den Seiten der *Politischen Vierteljahresschrift* berichteten Forschungsarbeiten weitere Bemühungen zur Aufstellung aussagekräftiger Theorien über politische und soziale Phänomene erleichtern und fördern werden. Alle Materialien, die diesen Prozess erleichtern, sind willkommene Elemente eines Datensatzes auf einem etablierten Datenrepository.