

Gemeinsame Gastforschungsregeln FDZ DESTATIS sowie FDSZ der Bundesbank

Prinzipien und Regeln

Technical Report 2023-04, Version: 2-0

Gültig ab 01. Juli 2023

DOI: 10.12757/75722.1

Deutsche Bundesbank, Forschungsdaten und Servicezentrum
Statistischen Ämter des Bundes und der Länder, Forschungsdatenzentren

Keywords: Gastforschungsregeln, Outputkontrolle, Publikationskontrolle, Datenschutz

Version: 2-0

DOI: [10.12757/75722.2](https://doi.org/10.12757/75722.2)

Citation: FDSZ Bundesbank und FDZ DESTATIS (2023). Gemeinsame Gastforschungsregeln FDZ DESTATIS sowie FDSZ der Bundesbank, Technical Report 2023-04 - Version 2-0

Inhaltsverzeichnis

| | | |
|----------|--|-----------|
| 1 | Einleitung | 1 |
| 2 | Regeln zur Arbeit in der gesicherten Umgebung des FDZ | 2 |
| 3 | Prinzipien zur Outputprüfung | 4 |
| | Prinzip O.1: Anonymität..... | 4 |
| | Prinzip O.2: Nachvollziehbarkeit..... | 6 |
| | Prinzip O.3: Reproduzierbarkeit..... | 7 |
| | Prinzip O.4: Ressourcensparsamkeit | 8 |
| | Prinzip O.5: Verantwortlichkeit | 9 |
| 4 | Prinzipien zur Publikationsprüfung..... | 10 |
| | Prinzip P.1: Kontrolle aller Publikationen durch das FDZ und FDSZ | 10 |
| | Prinzip P.2: Kriterien zur Veröffentlichung von Forschungsergebnissen | 11 |
| 5 | Glossar | 12 |

1 Einleitung

Das Forschungsdatenzentrum (FDZ) von Destatis sowie das Forschungsdaten- und Servicezentrum (FDSZ) der Bundesbank bieten hochwertige Services für den Datenzugang zu verknüpften Mikrodaten beider Institutionen. Einfache und schnelle Antragswege, eine kompetente Nutzerbetreuung und Datenveredelung sind das Kerngeschäft von FDZ und FDSZ. Dabei gehen FDZ und FDSZ vertrauensvoll mit den ihnen bereitgestellten Daten um und berücksichtigen alle gesetzlichen und internen Regelungen zur Datenverarbeitung und Datenweitergabe.

Gastforschende, die mit den verknüpften Mikrodaten arbeiten, sind dafür verantwortlich, dass ihre Berechnungsergebnisse nicht die Datenvertraulichkeit verletzen. Sollen ihre Forschungsergebnisse die sichere Umgebung des FDZ verlassen, haben sie dafür Sorge zu tragen, dass auf Basis ihrer Berechnungsergebnisse keine Rückschlüsse auf einzelne Merkmalsträger oder statistische Einheiten wie beispielsweise Banken, (nichtfinanzielle) Unternehmen, Personen oder Haushalte möglich sind.

Die Einhaltung dieser Anforderungen an die Datenvertraulichkeit bei der Erstellung dieser Ergebnisse wird zusätzlich von Mitarbeitern des FDZ und FDSZ geprüft. Sollten diese Anforderungen nicht erfüllt sein, so können die Berechnungsergebnisse nicht herausgegeben bzw. Publikationen nicht freigegeben werden. Bei besonders starken Verstößen können darüber hinaus weitere, im Nutzungsvertrag genannte Maßnahmen ergriffen werden.

Dieses Dokument soll Gastforschenden dabei helfen, die vorgeschriebenen Anforderungen an den Gastforschungsaufenthalt, die Herausgabe von Berechnungsergebnissen sowie die Freigabe der darauf beruhenden Publikationen leichter zu erfüllen. Darüber hinaus stärken sie das Vertrauen der Datengeber, wie Banken und Unternehmen, und der Datenproduzenten, wie den Fachbereichen der Bundesbank, in den sicheren Umgang mit ihren Daten.

Im Falle von Unklarheiten oder Rückfragen können sich Gastforschende jederzeit an das FDZ (forschungsdatenzentrum@destatis.de) oder das FDSZ (fdsz-data@bundesbank.de) wenden. Diesbezüglich sollten alle Unklarheiten vor dem ersten Gastaufenthalt geklärt sein.

FDZ und FDSZ behalten sich vor, die im folgenden genannten Prinzipien und Regeln bei Bedarf anzupassen, zu ergänzen oder zu erweitern.

Das folgende Dokument enthält Prinzipien und Regeln zur (i) Arbeit in der gesicherten Umgebung des FDSZ, (ii) Outputprüfung und (iii) Publikationsprüfung. Darüber hinaus enthält das letzte Kapitel ein Glossar mit vielen Definitionen zu den in diesem Dokument verwendeten Begriffen.

2 Regeln zur Arbeit in der gesicherten Umgebung des FDZ

Beim Umgang mit Mikrodaten im Rahmen eines wissenschaftlichen Forschungsprojektes sind Gastforschende verpflichtet, das Datengeheimnis zu wahren. Um die Kriterien der Anonymität in der gesicherten Umgebung des FDZ zu gewährleisten, müssen alle Gastforschenden die folgenden Regeln beachten:

1. Datenzugang nur mit Genehmigung. Daten dürfen im Rahmen eines Forschungsprojektes nur von zugelassenen Gastforschenden genutzt werden. Zugelassene Gastforschende haben mit beiden Institutionen, dem FDZ und FDSZ, einen entsprechenden Vertrag unterzeichnet und wurden beim ersten Gastforschungsaufenthalt verpflichtet. Die Einzeldaten dürfen Dritten nicht offenbart werden. Als Dritte gelten auch Ko-Autoren/innen, die keine zugelassenen Gastforschenden sind.
2. Keine Nutzung externer Hardware. Die Nutzung von Laptops, Smartphones, Tablets und anderer elektronischer Geräte ist am Gastforschungsarbeitsplatz verboten.
3. Keine Kopien von Einzeldaten. Alle Daten verbleiben in dem für das Projekt vorgesehenen Speicherbereich. Sie dürfen nicht abgeschrieben, ausgedruckt, abfotografiert oder außerhalb des Projektordners (beispielsweise auf der lokalen Festplatte oder dem Desktop) abgespeichert werden.
4. Re-Identifikationsverbot. Für die Verwendung von Einzeldaten besteht ein absolutes Re-Identifikationsverbot. Die Gastforschenden haben eine De-Anonymisierung der ihnen zur Verfügung gestellten anonymisierten Einzeldaten zu unterlassen.
5. Einhaltung der vorgegebenen Ordnerstruktur. Gastforschende erhalten Zugriff auf einen spezifischen Projektordner, der mit ihrer Projektnummer gekennzeichnet ist. Die Verzeichnisstruktur innerhalb des Projektordners sieht eine Trennung von Gastwissenschafts-arbeitsplatz-Daten (.data\gwap), eventuell verwendeten externen Daten (.data\ext), notwendigen temporären Daten (.data\temp), Programmcodes (.code), Ergebnissen (.results) und Ergebnisse für die Outputprüfung (.transfer) vor. Diese vom FDZ und FDSZ vorgegebene Ordnerstruktur (inklusive Unterordner) darf nicht verändert oder gelöscht werden. Innerhalb des spezifischen Projektordners dürfen, entsprechend der vorgegebenen Regeln, eigene Ordner erstellt werden. Die maximale Pfadlänge für Dateinamen inklusive der dazugehörigen Pfadstruktur beträgt 130 Zeichen.
6. Berechtigung zur Nutzung externer Daten. Die Nutzung DESTATIS- sowie Bundesbank-externer Daten im Rahmen eines Forschungsprojektes erfordert eine Bestätigung der Datennutzungsberechtigung, die vor Projektbeginn schriftlich beim FDSZ eingereicht werden muss. Im FDZ muss für externe Datensätze eine Verfahrensbeschreibung der durchzuführenden Verknüpfung eingereicht werden.

7. Rechtzeitige Übermittlung externer Daten. Externe Daten, die an einem Gastforschungsarbeitsplatz genutzt werden sollen, müssen mindestens fünf Werktage vor dem geplanten Gastforschungsaufenthalt, per E-Mail zur Prüfung an das FDZ übermittelt werden.
8. Rechtzeitige Übermittlung externer Programmcodes. Vorher angefertigte Programmcodes, die an einem Gastforschungsarbeitsplatz genutzt werden sollen, müssen mindestens vier Werktage vor dem geplanten Gastforschungsaufenthalt per E-Mail zur Prüfung an das FDZ übermittelt werden.
9. Nutzung genehmigter Software. Am Gastforschungsarbeitsplatz steht standardmäßig Stata zur Verfügung. Mit Microsoft Excel erzeugte oder in einem Excel-Dokument gespeicherte Berechnungsergebnisse werden nicht zur Outputprüfung akzeptiert und nicht freigegeben.
10. Keine Modifikationen an bereitgestellter Soft- und Hardware. Unbefugte Systemmodifikationen, z. B. das Deaktivieren von Schutzfunktionen, und das Einrichten lokaler Administrationsrechte für im Netz der Bank betriebene PC-Arbeitsplätze sind unzulässig. Eigenmächtiges Installieren von Software auf bankeigenen Geräten sowie die Nutzung von nicht freigegebener/privater Software (auch sog. portable Versionen) auf bankeigenen Geräten ist untersagt. Gleiches gilt für die Infrastruktur in DESTATIS und der Bundesbank im Allgemeinen.
11. Beachtung der Projektlaufzeit. Eine Nutzung der für das Projekt bereitgestellten Daten ist nur während der Projektlaufzeit zulässig. Für Gastforschende endet die Projektlaufzeit mit dem Ende der Vertragslaufzeit.

3 Prinzipien zur Outputprüfung

Die folgenden fünf Prinzipien und dazugehörigen Regeln sollen Gastforschende dabei unterstützen, die Regeln für die Prüfung von Forschungsergebnissen auf Geheimhaltung und Datenvertraulichkeit („Outputprüfung“) leichter zu erfüllen.

Prinzip O.1: Anonymität

Alle Berechnungsergebnisse, die die gesicherte Umgebung des FDZ und FDSZ verlassen, müssen absolut anonym sein. Absolut anonyme Ergebnisse lassen weder durch eine direkte Identifikation der Berichtspflichtigen oder sonstigen juristischen oder natürlichen Personen, Rechtssubjekte oder Niederlassungen (z.B. durch ihren Namen oder ihre Anschrift oder einen offiziell vergebenen Erkennungscode) noch indirekt durch Ableitung eine Re-Identifikation zu. Hierbei sind sämtliche Möglichkeiten zu berücksichtigen, die von einem Dritten in angemessener Form zur Identifizierung der Einzeldaten genutzt werden könnten. Gastforschende sind dafür verantwortlich, dass ihre Berechnungsergebnisse diese Bedingungen erfüllen. Zur Sicherstellung der absoluten Anonymität müssen folgende Regeln erfüllt sein:

1. Keine Herausgabe von Identifikatoren. Weder in den Berechnungsergebnissen noch in Programmcodes dürfen Identifikatoren enthalten sein.
2. Keine Herausgabe von Einzeldaten. Forschungsergebnisse dürfen keine Einzeldaten enthalten. Dies gilt sowohl für die Herausgabe von Datensätzen, Tabellen oder Grafiken, als auch für Befehle und Kommentare in Programmcodes (z.B. Stata Do-Files) und Log Files (z.B. Stata Log-Files). Minima und Maxima sind i.d.R. Einzeldaten. Quantile werden auch als Einzeldaten betrachtet, sollten sie nicht die Anforderungen unter Regel 5: Berechnung von Quantilen erfüllen.
3. Einhaltung der Mindestfallzahl. Herauszugebende Ergebnisse müssen auf mindestens sechs unterschiedlichen Merkmalsträgern basieren.
4. Einhaltung der p%- bzw. Dominanzregel. Wenn ein Aggregat aus fünf oder mehr Merkmalsträgern ausgewiesen wird, besteht weiterhin die Gefahr einer indirekten Identifizierung großer Merkmalsträger, falls diese zusammen einen ausreichend hohen Anteil zu diesem Wert beisteuern. Daher dürfen die Anteile der zwei größten Merkmalsträger am Gesamtwert zusammen einen gewissen Prozentwert nicht übersteigen¹.
5. Berechnung von Quantilen. Werden Quantile q einer Grundgesamtheit oder Untergruppe berechnet, so müssen im Bereich zwischen dem größten Quantil und dem Maximum mindestens sechs unterschiedliche Merkmalsträger vorliegen. Aus Symmetriegründen wird dies auch für den Bereich zwischen dem Minimum und dem kleinsten

¹ Diese bezieht sich auf nicht-transformierte Werte.

Quantil gefordert. Das Dominanzkriterium muss ebenfalls eingehalten werden. Liegen im Intervall von Quantil q_p bis Quantil $q_{(p+1)}$ weniger als sechs unterschiedliche Merkmalsträger, kann keine Freigabe der entsprechenden Quantile erfolgen.

6. Tabellenübergreifende Geheimhaltung. Sofern Berechnungsergebnisse zunächst basierend auf der Grundgesamtheit G und anschließend auf einer Teilmenge $X \subset G$ ausgegeben werden, muss die Einhaltung der oben genannten Regeln auch für die Differenz nachgewiesen werden. Andernfalls könnte eine Identifikation einzelner Merkmalsträger durch diese Differenzbildung nicht ausgeschlossen werden.
7. Behandlung von Dummy-Variablen. Wird der Mittelwert einer mit 0 und 1 kodierten Dummy-Variablen ausgegeben, so müssen mindestens sechs unterschiedliche Merkmalsträger die Ausprägung 0 haben und mindestens sechs unterschiedliche Merkmalsträger die Ausprägung 1 aufweisen.
8. Behandlung von Nullen und fehlenden Werten. Bei dichotomen und kategorialen Variablen sind Nullen bei deskriptiven Statistiken und bei Regressionsanalysen zulässige Werte, solange sie nicht als Code für fehlende Werte verwendet werden. Im Fall stetiger Variablen ist die Verwendung von Nullen in Regressionsanalysen erlaubt. Bei deskriptiven Statistiken dürfen Nullen und fehlende Werte bei der Ermittlung der Anzahl unterschiedlicher Merkmalsträger nicht mitgezählt werden.
9. Behandlung von Regionalinformationen. Deskriptive Analysen auf regionaler Ebene sind nicht erlaubt, wenn die regionalen Einheiten kleinräumiger als Kreise (NUTS3) sind. Bei Analysen auf regionaler Ebene bzw. für bestimmte regionale Einheiten, müssen die Ergebnisse auf mindestens 20 Merkmalsträgern basieren und das Dominanzkriterium muss erfüllt sein. In Regressionen können auch Regionaldummies für kleinräumigere regionale Einheiten verwendet werden, beispielsweise um für regionsspezifische Effekte zu kontrollieren. Bitte beachten Sie, dass die Koeffizienten für die Regionaldummies grundsätzlich nicht herausgegeben werden. Darüber hinaus dürfen Regionalinformationen keine indirekte Identifikation der beobachteten Einheit ermöglichen.

Für bestimmte Datensätze gelten besondere Geheimhaltungsvorschriften. Daher sind bei der Arbeit mit solchen Daten folgende zusätzliche Kriterien zu beachten:

- i. Mikrodatenbank Direktinvestitionen (MiDi). Bei der Ermittlung der Fallzahlen ist darauf zu achten, dass es nicht zu einer Mehrfachzählung kommt und jedes Unternehmen nur einmal gezählt wird, auch wenn es mehrere Investitionen tätigt. Übergeordnete Unternehmen sind ebenfalls als zu schützende Einheit zu betrachten. Im Falle eines Outward-Investments ist die zu schützende Einheit die nu2 (das Direktinvestitionsobjekt im Ausland / die ausländische Tochter), die num (deutsche Melder), sowie ggfs. die nui (deutsche Mutter des Melders), falls mehrere num's in einem Aggregat zu einer nui

gehören. Bei einem Inward-Investment sind entsprechend zu schützen: num, nu4 (ausländische Mutter), noa (ggf. Konzernmutter der ausländischen Mutter).

- ii. Statistik zum internationalen Dienstleistungshandel (SITS). Der Datensatz SITS bezieht sich auf Dienstleistungstransaktionen. D.h. jede Zeile bildet eine für ein bestimmtes Datum (Jahr und Monat) gemeldete Transaktion ab. Da ein Unternehmen in der Regel mehrere Transaktionen tätigt, können dadurch mehrere Zeilen Beobachtungen zu ein und demselben Unternehmen beinhalten. Die Anzahl der Zeilen mit ein und demselben Monat entspricht daher nicht der Zahl der Unternehmen. Bei der Ermittlung der Fallzahlen ist darauf zu achten, dass es nicht zu Mehrfachzählungen kommt, sondern jedes Unternehmen nur einmal gezählt wird, auch wenn es mehrere Transaktionen tätigt.

Prinzip O.2: Nachvollziehbarkeit

Zur Überprüfung der Einhaltung der in Prinzip O.1 genannten Anforderungen an die Anonymität, muss es Mitarbeitende des FDZ und FDSZ möglich sein, die Erstellung der Forschungsergebnisse mit vertretbarem Zeitaufwand nachzuvollziehen. Aus diesem Grund sind Gastforschende verpflichtet folgende Regeln einzuhalten:

1. Master-File. Eine Master-Datei muss angelegt werden, die alle relevanten Informationen zum Forschungsprojekt enthält und alle verwendeten Unterprogramme aufruft. Stata-Templates ("0_master.do", "1_data_preparation.do", "2_descriptive_analysis.do" und "3_regressions.do") werden vom FDZ und FDSZ zur Verfügung gestellt.
2. Logging. Für jeden Programmcode muss die Protokollfunktion (log-Datei) aktiviert werden. Die Protokollierung muss bereits vor der ersten Anweisung und vor der inhaltlichen Beschreibung des Forschungsprojekts erfolgen.
3. Ordnung und Struktur im Programmcode. Der Programmcode ist optisch klar zu strukturieren, sodass einzelne Programmblöcke (Programmkopf, einzelne Auswertungsschritte usw.) optisch voneinander abgegrenzt sind. Schleifen müssen eingerückt sein. Lange Programme bzw. Analyseschritte müssen auf kleinere Programmdateien aufgeteilt werden, z.B. Datenaufbereitung, deskriptive Auswertungen, Regressionen (siehe Stata-Templates unter Regel 2: Logging).
4. Kommentierung des Programmcodes. Der Programmcode muss gut kommentiert sein, sodass auch eine nicht mit dem Projekt vertraute Person diesen mit vertretbarem Zeitaufwand nachvollziehen kann.
5. Klare Benennung von Ergebnisdateien. Sämtliche Ergebnisdateien müssen mit dem gleichen Namen beginnen wie das Programm, mit dem diese Dateien generiert werden, und nachvollziehend nummeriert werden.

6. Klare Benennung von Variablen. Alle vergebenen Namen müssen möglichst sprechend sein und einheitlich verwendet werden. Für alle selbst erzeugten und alle extern eingebrachten Daten ist die Vergabe von Variablenlabels sowie eine kurze Variablenbeschreibung erforderlich. Werden (kategoriale) Variablen neu erzeugt oder verändert, dann müssen für die Ausprägungen entsprechende Wertelabels vergeben werden.
7. Angabe von Fallzahlen. Die Anzahl der den Berechnungen zugrundeliegenden Merkmalsträger muss für jegliche Form der Ergebnisdarstellung (z.B. Tabellen, Grafiken, Auswertungen) angegeben werden. In Stata können die Fallzahlen mit den vom FDZ und FDSZ bereitgestellten Befehlen `nobsdes6` und `nobsreg6` ausgegeben und überprüft werden. Die entsprechenden ado-Files stehen in der FDZ- und FDSZ-`ado_library` und auf der FDSZ-Homepage zur Verfügung.
8. Wiederholte Prüfung. Bei geringfügigen Änderungen der Programmcodes und erneuter Einreichung des Outputs ist explizit auf die getätigten Änderungen hinzuweisen. Sofern möglich lassen die Gastforschende nur die betroffenen Programmelemente noch einmal prüfen.

Prinzip O.3: Reproduzierbarkeit

Um die Einhaltung der in Prinzip O.1 genannten Anforderungen an Anonymität überprüfen zu können, müssen alle zur Prüfung vorgelegten Berechnungsergebnisse reproduzierbar sein. Gastforschende sind verpflichtet, die folgenden Regeln einzuhalten:

1. Reproduzierbarkeit des Programmcodes. Alle Berechnungsergebnisse müssen lückenlos und von einem fehlerfrei lauffähigen Programm mit dem Namen „0_master.do“ erzeugt werden, welches alle im Projektverlauf relevanten Auswertungsprogramme enthalten muss. Dieses Programm muss zwingend mit dem Laden der vom FDZ und FDSZ zur Verfügung gestellten Originaldaten beginnen. Das Programm muss bezüglich der schon einmal zur Überprüfung vorgelegten Ergebnisse immer dieselben Verarbeitungsschritte durchlaufen und exakt dieselben Ergebnisse erzeugen. In der „0_master.do“-Datei muss hinter jedem Programmaufruf eine kurze Beschreibung der Inhalte des jeweiligen Unterprogramms erfolgen.
2. Reproduzierbarkeit der Software. Jegliche zur Erzeugung der Berechnungsergebnisse genutzte Software muss zu Beginn der „0_master.do“-Datei nachvollziehbar genannt werden (Name und Version). Neben der Benennung der Version der Analysesoftware schließt das auch das Benennen aller zusätzlich genutzten Pakete (z.B. R, Python, Matlab) oder von anderen Nutzenden generierten Programmcodes (z.B. Stata ado-files) und deren jeweilige Version mit ein. Dies gilt auch für vom FDZ / FDSZ bereitgestellte Software.

3. Reproduzierbarkeit der Daten. Alle zur Erzeugung der Berechnungsergebnisse genutzten Datensätze des FDZ und FDSZ müssen zu Beginn der „0_master.do“-Datei eindeutig und nachvollziehbar genannt werden (DOI, falls vorhanden, Statistik und Jahr). Die genutzten Datensätze, die von externen Datenanbietern stammen, müssen vor der Nutzung im Forschungsprojekt zusammen mit einer geeigneten Dokumentation der Daten und der Quellen eingeschickt werden.
4. Reproduzierbarkeit der Veröffentlichung. Alle veröffentlichten Berechnungsergebnisse müssen schnell, einfach und eindeutig in dem durch die Auswertungsprogramme erzeugten Output gefunden werden. Hierzu ist es notwendig, dass eine eindeutige Beziehung zwischen Auswertungsprogramm, Output und Publikation hergestellt wird. Dies kann zum Beispiel so geschehen, dass der Forscher seiner Publikation ein Dokument beifügt, das zu jedem Berechnungsergebnis den Bezug zum Tabellennamen im Output herstellt. Neben Tabellen und Graphiken sind auch alle Berechnungsergebnisse im Text in diesem Dokument aufzuführen. Das FDSZ empfiehlt ein „master_yyyymmdd_publication.do“ zu erzeugen, in dem alle Berechnungsergebnisse mit Verweisen auf die entsprechenden Berechnungen, Tabellen oder Grafiken in der Publikation enthalten sind. Sollte die Reproduzierbarkeit der Veröffentlichung nicht hergestellt sein, so kann keine Freigabe der Publikation erfolgen.

Prinzip O.4: Ressourcensparsamkeit

Ergebnisse, die die gesicherte Umgebung des FDZ oder FDSZ verlassen, sollen grundsätzlich dem Zwecke einer Publikation dienen. Daher sollen Gastforschende bei der Entscheidung, welche Ergebnisse zur Prüfung vorzulegen sind, Ressourcensparsamkeit walten lassen. Die vorgelegten Berechnungsergebnisse sind auf eine Anzahl zu beschränken, welche üblicherweise in einem empirischen wissenschaftlichen Artikel Platz findet. Zudem sollten Gastforschende bei der Entscheidung, welche Ergebnisse zur Prüfung vorgelegt werden, berücksichtigen, dass insbesondere die Prüfung der „tabellenübergreifenden Geheimhaltung“ (siehe Prinzip O.1 – Regel 6) einen erheblichen zeitlichen Aufwand erfordert. Grundsätzlich müssen Gastforschende folgende Regeln beachten:

1. Keine explorativen Analysen. Nur potenziell publikationsfähige Analysen dürfen zur Prüfung vorgelegt werden. Von den zahlenmäßig umfangreichen Berechnungsergebnissen, wie sie häufig im Rahmen explorativer Analysen oder im Rahmen der Prüfung verschiedener Modellspezifikationen bei Regressionen anfallen, sind bereits während des Forschungsaufenthalts im FDZ oder FDSZ die publikationswürdigen Varianten auszuwählen.
2. Maximale Zeilenanzahl für Output. Im Laufe eines Forschungsprojekts darf die Anzahl der insgesamt zur Prüfung vorgelegten Zeilen die maximale Anzahl von 2.500 grundsätzlich nicht überschreiten. Freizugebende Grafiken werden in der Regel pauschal mit einer Zeilenanzahl von 25 angesetzt. Da reine Code-Files (z.B. Stata „Do-Files“) keine

Ergebnisse enthalten dürfen, werden diese bei der Zählung nicht berücksichtigt. Das FDZ / FDSZ benachrichtigt die Gastforschenden sobald eine Zeilenanzahl von 2.000 überschritten wurde.

3. Erhöhung der maximalen Zeilenanzahl für Output. Gastforschende können in äußersten Ausnahmefällen einmalig eine Erhöhung der maximalen Zeilenanzahl beantragen. Hierfür muss eine ausführliche Begründung vorgelegt werden, warum zusätzlicher Output benötigt wird. Hierbei haben Gastforschende ausführlich darzustellen,
 - a. warum die maximale Zeilenanzahl nicht ausreicht,
 - b. wie die Tabellenübergreifende Geheimhaltung (siehe Prinzip O.1 – Regel 6) gewährleistet werden soll,
 - c. wie viele zusätzliche Zeilen bis zum Projektende noch benötigt werden,
 - d. warum diese zusätzlichen Zeilen bis zum Projektende noch benötigt werden (ausführliche Beschreibung der geplanten Aktivitäten mit Mengenabschätzung).
 - e. ausführliche Erläuterung, welche projektspezifischen Maßnahmen ergriffen werden, damit die neu festzulegende Grenze eingehalten wird.

Auf Grundlage dieser Begründung legt das FDZ und FDSZ die neue Zeilenzahl für den Output fest. Sollte die Begründung nicht ausreichend sein, wird der Antrag auf Erhöhung der maximalen Zeilenanzahl für den Output vom FDZ und FDSZ mit entsprechender Begründung abgelehnt. Eine einmalige Nachbesserung nach Ablehnung ist möglich.

4. Maximale Zeilenanzahl für Output bei Projekten, an denen Forschende der Deutschen Bundesbank beteiligt sind. Nicht gezählt wird der Output, wenn Forschende der Deutschen Bundesbank die Prüfung des Outputs selbst vornehmen. Diese Ausnahme gilt nicht für Projekte, bei denen das FDZ und FDSZ die Outputprüfung durchführt bzw. durchführen muss.
5. (IT-)ressourcenbewusste Programmierweise. Gastforschende sollen bei der Entscheidung, welche umfangreicheren Berechnungen für das Forschungsprojekt wirklich notwendig sind, sparsam vorgehen. Dies betrifft Laufzeiten, Rechnerlast und generiertes Datenvolumen.

Prinzip O.5: Verantwortlichkeit

Gastforschende sind für die Einhaltung aller in diesem Dokument aufgeführten Prinzipien und Regeln verantwortlich. Hierüber werden alle Gastforschende vor Beginn des Forschungsprojektes im Rahmen einer Verpflichtung detailliert aufgeklärt. Eine Nichteinhaltung führt zur Verweigerung der Herausgabe der Berechnungsergebnisse seitens des FDZ / FDSZ. Dabei müssen Gastforschende die folgenden Regeln beachten:

1. Kontrolle aller herauszugebenden Berechnungsergebnisse. Bevor die Gastforschenden Berechnungsergebnisse, außerhalb der gesicherten Umgebung verwenden können, muss zunächst durch das FDZ / FDSZ kontrolliert werden, ob die Gastforschenden bei ihren Berechnungsergebnissen die Prinzipien zur Outputprüfung angewendet haben, und die Berechnungsergebnisse müssen vom FDZ und FDSZ freigegeben werden. Gastforschende müssen das FDZ / FDSZ daher per E-Mail darüber informieren, welche Berechnungsergebnisse geprüft werden sollen. Diese müssen stets in einem mit dem entsprechenden Datum benannten Unterordner von "transfer" abgelegt werden. Zusätzlich zur Prüfung notwendige Informationen, die jedoch nicht herausgegeben werden müssen, müssen in einem mit dem gleichen Datum benannten Unterordner von "results" auffindbar sein. Die Versendung von Berechnungsergebnissen durch die Datennutzenden als Anhang in einer Email ist nicht erlaubt. Die inhaltliche Sinnhaftigkeit der übermittelten Forschungsergebnisse wird nicht geprüft.
2. Funktionsfähige Programmcodes. Enthält der Programmcode syntaktische oder sonstige Fehler, werden diese vom FDZ / FDSZ nicht behoben und die Gastforschende werden zur Nachbesserung des Programmcodes aufgefordert.
3. Format für Outputprüfung. Ergebnisse und Programmcodes werden nur zur Outputprüfung akzeptiert, wenn diese editierbar sind und als „plain text files“ vorliegen. Grafiken müssen in einem nicht weiterzuverarbeitenden Format vorliegen.

4 Prinzipien zur Publikationsprüfung

Die folgenden zwei Prinzipien und dazugehörigen Regeln sollen Gastforschende dabei unterstützen, die Regeln für die Prüfung von Publikationen („Publikationsprüfung“) leichter zu erfüllen.

Prinzip P.1: Kontrolle aller Publikationen durch das FDZ und FDSZ

1. Kontrolle aller Publikationen durch das FDZ und FDSZ. Forschungsergebnisse, die Berechnungsergebnisse beinhalten, dürfen erst veröffentlicht werden, wenn eine schriftliche Zustimmung der prüfenden Stellen vorliegt. Die Einwilligung kann verweigert werden, wenn die zur Veröffentlichung bestimmten Ergebnisse, den in Prinzip P.2 genannten Kriterien nicht entsprechen. Die Kontrolle der Publikation kann bis zu 10 Werktagen dauern.
2. Meldepflicht. Gastforschende sind dafür verantwortlich jede angestrebte Veröffentlichung, die Forschungsergebnisse aus den Auswertungen während des Gastforscheraufenthaltes im FDZ / FDSZ beinhaltet, dem FDZ / FDSZ vorzulegen.
3. Bezug auf zuvor freigegebener Berechnungsergebnisse / Reproduzierbarkeit der publizierten Ergebnisse. Publikationen dürfen ausschließlich nur durch das FDZ und FDSZ

freigegebene Berechnungsergebnisse enthalten. Bei Einreichen einer Publikation zur Prüfung müssen Gastforschende die Anforderungen in Prinzip O.3, Regel: Reproduzierbarkeit der Veröffentlichung erfüllen. Sollte dies nicht der Fall sein, so kann die Publikation nicht zur Veröffentlichung freigegeben werden.

4. Mehrfachveröffentlichungen. Für weitere Veröffentlichungen, die ausschließlich auf bereits genehmigten Veröffentlichungen basieren, ist eine erneute Einwilligung der einwilligenden Stelle nicht notwendig. Werden Berechnungsergebnisse verwendet, die zuvor nicht in einer bereits genehmigten Publikation enthalten waren, so müssen diese kenntlich gemacht und die Publikation zur Prüfung an das FDZ / FDSZ gegeben werden.

Prinzip P.2: Kriterien zur Veröffentlichung von Forschungsergebnissen

Eine Veröffentlichung muss darüber hinaus die folgenden Kriterien erfüllen:

1. Benennung des Forschungsprojektes. Jede Publikation muss die Projektnummer nennen.
2. Benennung der Art des Datenzugangs. In jeder Publikation muss die Art des Datenzugangs, d.h. Gastforscheraufenthalt (GaFo)/Gastwissenschaftsarbeitsplatz (GWAP), Kontrollierte Datenfernverarbeitung (KDFV) oder Scientific Use File (SUF), benannt werden.
3. Benennung der verwendeten Datensätze. Alle im Forschungsprojekt verwendeten Datensätze müssen mit Namen und wenn vorhanden mit DOI und darüber hinaus der zugehörigen Datendokumentation oder entsprechender Publikationen zitiert werden.
4. Zitation. Jede grafische oder tabellarische Darstellung ist mit einem Zitierhinweis zu versehen: Quelle: Forschungsdaten- und Servicezentrum (FDSZ) der Deutschen Bundesbank und Forschungsdatenzentren (FDZ) des Bundes und der Länder , <Name des genutzten Einzeldatensatzes aus dem Standardangebot des FDSZ/FDZ (gegebenenfalls einheitlich verwendete Abkürzung)>, <Verwendungszeitraum des genutzten Einzeldatensatzes>, <DOI>, eigene Berechnungen.

5 Glossar

| Begriff | Erklärung |
|-----------------------|--|
| Absolute Anonymität | Absolut anonymisierte Ergebnisse lassen sich weder durch eine direkte Identifikation von Berichtspflichtigen oder sonstigen juristischen oder natürlichen Personen, Rechtssubjekten oder Niederlassungen (z.B. durch ihren Namen oder ihre Anschrift oder einen offiziell vergebenen Erkennungskode) oder indirekt durch Ableitung identifizieren. Hierbei sind sämtliche Möglichkeiten zu berücksichtigen, die von einem Dritten in angemessener Form zur Identifizierung der Einzeldaten genutzt werden könnten. |
| Aggregierte Daten | Aggregierte Daten fassen eine Vielzahl an Einzeldaten in einem einzigen Wert zusammen. |
| Berechnungsergebnisse | Berechnungsergebnisse sind sämtliche im Rahmen der Durchführung des Forschungsprojekts entstandenen Werke, insbesondere Projektskizzen, Datenauswertungen, wissenschaftliche Abhandlungen und Präsentationen, auch soweit sie sich noch im Entwurfsstadium befinden, sowie sämtliche im Rahmen der Durchführung des Forschungsprojekts gewonnenen wissenschaftlichen Erkenntnisse und Know-how. |
| Datendokumentation | Das FDZ / FDSZ stellt für fast alle der bereitgestellten Mikrodatensätze Datendokumentationen mit Metadaten bereit. Diese können auf der Website des FDZ und FDSZ heruntergeladen werden. |
| Datengeheimnis | Mit der Datenverarbeitung befasste Personen dürfen personenbezogene Daten nicht unbefugt verarbeiten (Datengeheimnis). Sie sind bei der Aufnahme ihrer Tätigkeit auf das Datengeheimnis zu verpflichten. Das Datengeheimnis besteht auch nach der Beendigung ihrer Tätigkeit fort (vgl. § 53 des Bundesdatenschutzgesetzes (BDSG)). |
| Datenvertraulichkeit | Der Grundsatz der Integrität und Vertraulichkeit (Art. 5 Abs. 1 f. DSGVO) sieht vor, dass personenbezogene Daten in einer Art und Weise verarbeitet werden, die eine angemessene Sicherheit dieser Daten gewährleistet. Insbesondere ist dafür zu sorgen, dass die personenbezogenen Daten <ol style="list-style-type: none">1. vor unbefugter und unrechtmäßiger Verarbeitung (Vertraulichkeit) und2. vor unbeabsichtigtem Verlust, unbeabsichtigter Zerstörung oder unbeabsichtigter Schädigung (Integrität) bewahrt werden. Um dies zu erreichen, müssen geeignete technische und organisatorische Maßnahmen (TOM) getroffen werden. |

| | |
|---------------------------------|---|
| De-Anonymisierung | De-Anonymisierung ist die gezielte Aufhebung vorher durchgeführter Anonymisierung von Daten zur Re-Identifikation der zugrundeliegenden Einzeldaten. |
| Deskriptive Analyse | Mittels deskriptiver Analyse werden empirische Daten als Tabelle, Kennzahl oder Grafiken beschrieben und übersichtlich dargestellt. |
| Digital Object Identifier (DOI) | Der Digital Object Identifier (DOI) nach ISO 26324 ist ein eindeutiger und dauerhafter digitaler Identifikator für physische, digitale oder abstrakte Objekte. Alle vom FDZ / FDSZ bereitgestellten Standarddatensätze sind mit einem DOI registriert. |
| Direkte Identifikation | Identifikation von Berichtspflichtigen oder sonstigen juristischen oder natürlichen Personen, Rechtssubjekten oder Niederlassungen durch ihren Namen, ihre Anschrift oder einen offiziell vergebenen Erkennungscode, sodass Einzelangaben bekannt werden. |
| Dominanzkriterium | Merkmalsträger unterscheiden sich hinsichtlich ihrer Größe. Auch wenn ein Wert basierend auf fünf oder mehr Merkmalsträgern ausgewiesen wird, dann besteht die Gefahr einer indirekten Identifizierung großer Merkmalsträger, falls diese zusammen einen ausreichend hohen Anteil zu diesem Wert beisteuern (Dominanzkriterium). Die Möglichkeit einer solchen indirekten Identifizierung muss ausgeschlossen sein. Deshalb muss nachgewiesen werden, dass unter den Merkmalsträgern, die der Berechnung eines Werts zugrunde liegen, keine zwei Merkmalsträger zu finden sind, die im Vergleich zu den übrigen Merkmalsträgern deutlich größer sind. Als Faustregel gilt: Die Anteile der zwei größten Merkmalsträger am Gesamtwert sollten zusammen einen gewissen Prozentwert X nicht übersteigen. |
| Dritte | Dritte sind alle Personen, die keine Gastforschende oder Beschäftigte des FDZ / FDSZs sind. Hierzu zählen auch Ko-Autoren/innen ohne direkten Datenzugang. |
| Einzeldaten | Mikrodaten oder Einzeldaten sind sehr granulare Daten, die sich unmittelbar auf Erhebungseinheiten beziehen. |
| Faktischen Anonymität | Informationen sind faktisch anonym, wenn sie nur mit unverhältnismäßig großem Aufwand an Zeit, Kosten und Arbeitskraft, den zugrundeliegenden Einzeldaten zugeordnet werden können. Die vom FDSZ in der gesicherten Umgebung bereitgestellten formal anonymisierten Daten gelten ebenfalls als faktisch anonymisiert (vgl. § 16 Art. 6.2 Bundesstatistikgesetz (BStatG)). |
| Fallzahl | Anzahl der einer Auswertung zugrundeliegenden Merkmalsträger. Hierbei sind die unterschiedlichen IDs und nicht etwa die gesamte Anzahl der Beobachtungen gemeint. |

| | |
|--|---|
| Formale Anonymität | Für die Herstellung der formalen Anonymität werden die direkten Identifikationen aus dem Datenmaterial entfernt, der weitere Merkmalsumfang bleibt dagegen erhalten. |
| Forschungsergebnisse | Siehe Definition von <i>Berechnungsergebnisse</i> . |
| Gastforschungsarbeitsplatz | Der Arbeitsplatz an dem Gastforschende in der gesicherten Umgebung des FDZ und FDSZ arbeiten können. |
| Gastforschende | Alle Gastforschende unterzeichnen einen Vertrag mit dem FDZ und FDSZ und müssen vor ihrem ersten Gastforscheraufenthalt verpflichtet sein. Im Rahmen ihres Forschungsprojektes haben Gastforschende direkten Zugang zu vertraulichen Mikrodaten von Destatis und der Deutschen Bundesbank. |
| Gastforschungsaufenthalt | Persönlicher Aufenthalt von Gastforschende in den Räumlichkeiten des FDZ / FDSZs. Gastforschende müssen an einem Gastforscherarbeitsplatz in der gesicherten Umgebung arbeiten. |
| Gesicherte Umgebung | In der gesicherten Umgebung können Gastforschende an einem vom FDZ bereitgestellten PC arbeiten, der weder über eine Internetverbindung noch über eine Druckfunktion verfügt. Die Nutzung persönlicher elektronischer Geräte wie Handy, Laptop oder Tablet ist in der gesicherten Umgebung strengstens untersagt. Es dürfen keine ungeprüften Ergebnisse die gesicherte Umgebung verlassen. |
| Identifikator | Ein Identifikator ist ein Kennzeichen zur eindeutigen Identifizierung eines bestimmten Objekts. |
| Indirekte Identifikation | Identifikation von Berichtspflichtigen oder sonstige juristische oder natürliche Personen, Rechtssubjekten oder Niederlassungen durch Ableitung, sodass Einzelangaben bekannt werden. |
| Kontrollierte Datenfernverarbeitung (KDFV) | Mithilfe der KDFV können Forschende die Basis-Daten analysieren, ohne dabei physisch direkt auf die Daten zugreifen zu können. Beispielsweise liegt das AFiD-Panel Außenhandelsstatistik am GWAP nur als Stichprobe vor. |
| Merkmalsträger | Die Merkmalsträger sind die „zu schützenden“ Einheiten, d.h. alle Berichtspflichtigen oder sonstige juristische oder natürliche Personen, Rechtssubjekten oder Niederlassungen. Dies können beispielsweise Banken, Unternehmen, private Haushalte, Kreditnehmer oder Kreditgeber sein. |
| Mikrodaten | Siehe Definition von <i>Einzeldaten</i> . |
| Mindestfallzahl | Die in diesem Dokument dargelegte Mindestfallzahl beträgt grundsätzlich sechs Merkmalsträger. Siehe Definition von <i>Fallzahl</i> . |
| Output | Siehe Definition von <i>Berechnungsergebnisse</i> . |
| Outputprüfung | Das FDZ / FDSZ überprüft die Einhaltung der in diesem Dokument unter 3. Outputprüfung dargelegten Prinzipien und Richtlinien durch die Gastforschende. |

| | |
|---------------------|--|
| Plain Text Files | Mit plain text werden Daten bezeichnet, die direkt unter Verwendung einer Zeichenkodierung in Text umgesetzt werden können. Beispiele sind .csv- oder .txt-Dateien. |
| Projektnummer | Das FDZ / FDSZ teilt jedem Forschungsprojekt zu Beginn eine achtstellige Projektnummer zu. Diese setzt sich aus dem Jahr (YYYY) sowie einer fortlaufenden Nummerierung (xxxx) zusammen: YYYY/xxxx. |
| Projektordner | Jedes Forschungsprojekt, das mindestens einen Gastforscheraufenthalt umfasst, hat einen eigenen Projektordner auf internen Laufwerk. Dieser ist mit der Projektnummer benannt und für Gastforschende im Rahmen eines Gastforscheraufenthaltes zugänglich. |
| Prüfende Stelle | Das FDZ DESTATISs sowie das FDSZ der Bundesbank. |
| Publikationsprüfung | Das FDZ / FDSZ überprüft die Einhaltung der in diesem Dokument unter 4. Publikationsprüfung dargelegten Prinzipien und Richtlinien durch die Gastforschende. |
| Re-Identifikation | Siehe Definition von <i>De-Anonymisierung</i> . |
| Veröffentlichung | Als Veröffentlichung gilt unter anderem die Darstellung der Forschungsergebnisse. <ul style="list-style-type: none"> - in Zeitschriftenartikeln, - in Sammelbandbeiträgen, - in Büchern, - in Working Paper Series sowie Diskussionspapieren von Institutionen oder Unternehmen, - in Dissertationen, - bei jeglicher Online-Stellung, insbesondere auf Wissenschaftsplattformen wie SSRN.com und RePEc.org, - in Konferenz-Webseiten oder persönlichen Webseiten wie etwa Google Pages etc., - in Mitarbeiterseiten von wissenschaftlichen Einrichtungen. - in Einreichungen bei Konferenzen in Form von Arbeitspapieren, - in allgemein zugänglichen Präsentationsfolien, in allgemein zugänglichen Bachelor- und Masterarbeiten. Allgemein zugänglich ist eine Darstellung dann, wenn sie technisch geeignet und bestimmt ist, der Allgemeinheit, d.h. einem individuell nicht bestimmbar Personenkreis, Informationen zu verschaffen |
| Verpflichtung | Vor dem ersten Gastforschungsaufenthalt oder am ersten Tag des Gastforschungsaufenthaltes werden Gastforschende von Mitarbeitenden des FDZ und des FDSZ über die Verpflichtung nach § 1 Abs. 1 des Gesetzes über die förmliche Verpflichtung nichtbeamteteter Personen (Verpflichtungsgesetz) vom 2. März 1974 und über die Verpflichtung auf das Datengeheimnis unterrichtet. |

Zu schützende Einheit

Siehe Definition von *Merkmalsträger*.