

# Gastforscherregeln im FDSZ

## Prinzipien und Regeln

Technical Report 2021-02, Version: 1-0

Gültig ab 01. März 2021

DOI: 10.12757/75721.1

Deutsche Bundesbank, Forschungsdaten und Servicezentrum

**Keywords:** Gastforscherregeln, Outputkontrolle, Publikationskontrolle, Datenschutz

**Version:** 1-0

**DOI:** 10.12757/75721.1

**Citation:** Forschungsdaten und Servicezentrum (2021). Gastforscherregeln im FDSZ, Technical Report 2021-02 - Version 1-0, Deutsche Bundesbank, Forschungsdaten und Servicezentrum.

## Contents

<b>1 Einleitung</b> . . . . .	<b>4</b>
<b>2 Regeln zur Arbeit in der gesicherten Umgebung des FDSZ</b> . . . . .	<b>5</b>
<b>3 Prinzipien zur Outputprüfung</b> . . . . .	<b>7</b>
Prinzip O.1: Anonymität . . . . .	7
Prinzip O.2: Nachvollziehbarkeit . . . . .	9
Prinzip O.3: Reproduzierbarkeit . . . . .	10
Prinzip O.4: Ressourcensparsamkeit . . . . .	11
Prinzip O.5: Verantwortlichkeit . . . . .	12
<b>4 Prinzipien zur Publikationsprüfung</b> . . . . .	<b>13</b>
Prinzip P.1: Kontrolle aller Publikationen durch das FDSZ . . . . .	13
Prinzip P.2: Kriterien zur Veröffentlichung von Forschungsergebnissen . . . . .	13
<b>5 Glossar</b> . . . . .	<b>14</b>

## 1 Einleitung

Das Forschungsdaten- und Servicezentrum (FDSZ) bietet hochwertige Services für den Datenzugang zu Mikrodaten der Deutschen Bundesbank. Einfache und schnelle Antragswege, eine kompetente Nutzerbetreuung und Datenveredelung sind das Kerngeschäft des FDSZ. Dabei geht das FDSZ vertrauensvoll mit den ihm bereitgestellten Daten um und berücksichtigt alle gesetzlichen und internen Regelungen zur Datenverarbeitung und Datenweitergabe.

Gastforscher/innen, die mit Mikrodaten der Bundesbank arbeiten, sind dafür verantwortlich, dass ihre Berechnungsergebnisse nicht die Datenvertraulichkeit verletzen. Sollen ihre Forschungsergebnisse die sichere Umgebung des FDSZ 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 FDSZ geprüft<sup>1</sup>. 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 Vertrag genannte Maßnahmen ergriffen werden.

Dieses Dokument soll Gastforscher/innen dabei helfen, die vorgeschriebenen Anforderungen an den Gastforscheraufenthalt, 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 Gastforscher/innen jederzeit an [fdsz-data@bundesbank.de](mailto:fdsz-data@bundesbank.de) wenden. Diesbezüglich sollten alle Unklarheiten vor dem ersten Gastaufenthalt geklärt sein.

Das FDSZ behält 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.

---

<sup>1</sup> Im Normalfall ist somit das FDSZ die prüfende Stelle. Im Einzelfall kann ein anderer Bereich der Bundesbank für die Prüfung zuständig sein.

## 2 Regeln zur Arbeit in der gesicherten Umgebung des FDSZ

Beim Umgang mit Mikrodaten im Rahmen eines wissenschaftlichen Forschungsprojektes sind Gastforscher/innen verpflichtet, das Datengeheimnis zu wahren. Um die Kriterien der *faktischen Anonymität*<sup>2)</sup> in der gesicherten Umgebung des FDSZ zu gewährleisten, müssen alle Gastforscher/innen die folgenden Regeln beachten:

1. **Datenzugang nur mit Genehmigung.** Daten dürfen im Rahmen eines Forschungsprojektes nur von zugelassenen Gastforscher/innen genutzt werden. Zugelassene Gastforscher/innen haben mit dem FDSZ einen entsprechenden Vertrag unterzeichnet und wurden beim ersten Gastforscheraufenthalt verpflichtet. Die Einzeldaten dürfen Dritten nicht offenbart werden. Als Dritte gelten auch Ko-Autoren/innen, die keine zugelassenen Gastforscher/innen sind.
2. **Keine Nutzung externer Hardware.** Die Nutzung von Laptops, Smartphones, Tablets und anderer elektronischer Geräte ist am Gastforscherarbeitsplatz 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 Gastforscher/innen haben eine De-Anonymisierung der ihnen zur Verfügung gestellten anonymisierten Einzeldaten zu unterlassen.
5. **Einhaltung der vorgegebenen Ordnerstruktur.** Gastforscher/innen erhalten Zugriff auf einen spezifischen Projektordner, der mit ihrer Projektnummer gekennzeichnet ist. Die Verzeichnisstruktur innerhalb des Projektordners sieht eine Trennung von Originaldaten (`.\data\orig`), eventuell verwendeten externen Daten (`.\data\ext`), notwendigen temporären Daten (`.\data\temp`), Programmcodes (`.\code`), Ergebnissen (`.\results`) und Ergebnisse für die Outputprüfung (`.\transfer`) vor. Diese vom 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 Bundesbank-externer Daten im Rahmen eines Forschungsprojektes erfordert eine Bestätigung der Datennutzungsberechtigung, die vor Projektbeginn schriftlich beim FDSZ eingereicht werden muss.
7. **Rechtzeitige Übermittlung externer Daten.** Externe Daten, die an einem Gastforscherarbeitsplatz genutzt werden sollen, müssen mindestens fünf Werktage vor dem geplanten Gastforscheraufenthalt, per E-Mail zur Prüfung an das FDSZ übermittelt werden.
8. **Rechtzeitige Übermittlung externer Programmcodes.** Vorher angefertigte Programmcodes, die an einem Gastforscherarbeitsplatz genutzt werden sollen, müssen mindestens zwei Werktage vor dem geplanten Gastforscheraufenthalt per E-Mail zur Prüfung an das FDSZ oder einen anderen zuständigen Bereich der Bundesbank übermittelt werden.
9. **Nutzung genehmigter Software.** Am Gastforscherarbeitsplatz steht standardmäßig Stata zur Verfügung. Die Nutzung von Matlab, Python und R ist prinzipiell möglich, muss aber vorab schriftlich mit dem FDSZ vereinbart werden. Mit Microsoft Excel erzeugte oder in einem Excel-Dokument gespeicherte Berechnungsergebnisse werden nicht zur Outputprüfung akzeptiert und nicht freigegeben.

---

2 Gesetz über die Statistik für Bundeszwecke (Bundesstatistikgesetz - BStatG) § 16 Art. 6.1: „Einzelangaben übermitteln, wenn die Einzelangaben nur mit einem unverhältnismäßig großen Aufwand an Zeit, Kosten und Arbeitskraft zugeordnet werden können (faktisch anonymisierte Einzelangaben), (...)“.

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.
11. **Beachtung der Projektlaufzeit.** Eine Nutzung der für das Projekt bereitgestellten Daten ist nur während der Projektlaufzeit zulässig. Für Gastforscher/innen endet die Projektlaufzeit mit dem Ende der Vertragslaufzeit.

### 3 Prinzipien zur Outputprüfung

Die folgenden fünf Prinzipien und dazugehörigen Regeln sollen Gastforscher/innen 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 FDSZ verlassen, müssen *absolut anonym* sein. Absolut anonyme Ergebnisse lassen sich 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 Erkennungskode) noch indirekt durch Ableitung identifizieren. Hierbei sind sämtliche Möglichkeiten zu berücksichtigen, die von einem Dritten in angemessener Form zur Identifizierung der Einzeldaten genutzt werden könnten. Gastforscher/innen 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 fünf 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 85 % nicht übersteigen<sup>3)</sup>.
5. **Berechnung von Quantilen.** Werden Quantile  $q$  einer Grundgesamtheit oder Untergruppe berechnet, so müssen im Bereich zwischen dem größten Quantil und dem Maximum mindestens fünf unterschiedliche Merkmalsträger vorliegen. Aus Symmetriegründen wird dies auch für den Bereich zwischen dem Minimum und dem kleinsten Quantil gefordert. Das Dominanzkriterium muss ebenfalls eingehalten werden. Liegen im Intervall von Quantil  $q_p$  bis Quantil  $q_{(p+1)}$  weniger als fünf unterschiedliche Merkmalsträger, müssen für mindestens einen dieser Merkmalsträger 15 % oder mehr des Betrags der relevanten Variablen außerhalb des Intervalls liegen.
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.

---

<sup>3</sup> Diese bezieht sich auf nicht-transformierte Werte.

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 fünf unterschiedliche Merkmalsträger die Ausprägung 0 haben und mindestens fünf 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. Bei Analysen auf regionaler Ebene mit Haushaltsdaten, beispielsweise dem Panel on Household Finances (PHF) oder dem Bundesbank Online Panel - Households (BOP-HH), ist eine Mindestfallzahl von fünf Merkmalsträgern ausreichend und das Dominanzkriterium muss erfüllt sein. In Regressionen können auch Regionaldummies für kleinräumigere regionalen 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. **Auslandsstatus der Banken (AUSTA).** Eine Bank (Mutter) kann mehrere rechtlich unselbstständige Zweigniederlassungen und/oder rechtlich selbstständige Tochterbanken unterhalten. Daher muss bei der Analyse auf der Ebene der Auslandsniederlassungen immer auch die Anzahl der zugehörigen Mütter geprüft werden.
- ii. **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).
- iii. **Millionenkreditevidenz.** Die Mindestfallzahl für Merkmalsträger muss hier sowohl für Kreditnehmer, als auch für Kreditgeber eingehalten werden. Liegen Informationen zu übergeordneten Kreditgebern (Kreditgeberkonzern) und Kreditnehmern (Kreditnehmereinheit) vor, so sind diese ebenfalls als zu schützende Einheit zu betrachten.
- iv. **Statistik über Wertpapierinvestments (SHS-Base plus).** Die schützenswerte Einheit ist das meldepflichtige Institut und nicht das Wertpapier. Kritisch sind Eigenbestände von meldepf-



lichtigen Instituten an eigenen Wertpapieren, d.h. eine Kombination von Deponentensektor 1221 / 1222 / 1223, ISIN und Bank, da hierbei die Möglichkeit besteht, anhand der ISIN den Wertpapieremittenten und somit das meldepflichtige Institut herauszufinden. Daher ist auch ein Zuspiel von Wertpapiernamen oder sonstigen Informationen, die es ermöglichen den Emittenten zu identifizieren, nicht erlaubt.

- v. **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.
- vi. **Unternehmensbilanzen aus dem Refinanzierungsgeschäft (Ustan).** Manche Unternehmen melden mehrere Abschlüsse in einem Jahr. Bei der Ermittlung der Fallzahlen ist daher darauf zu achten, dass solche Unternehmen nur einmal gezählt werden und es nicht zu Mehrfachzählungen kommt.

## Prinzip O.2: Nachvollziehbarkeit

Zur Überprüfung der Einhaltung der in *Prinzip O.1* genannten Anforderungen an die Anonymität, muss es Mitarbeiter/innen des FDSZ möglich sein, die Erstellung der Forschungsergebnisse mit vertretbarem Zeitaufwand nachzuvollziehen. Aus diesem Grund sind Gastforscher/innen 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 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 FDSZ bereitgestellten Befehlen *nobsdes5* und *nobsreg5* ausgegeben und überprüft werden. Die entsprechenden ado-Files stehen in der 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 Gastforscher/innen 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. Gastforscher/innen sind verpflichtet, die folgenden Regeln einzuhalten:

1. **Reproduzierbarkeit des Programmcodes.** Alle Berechnungsergebnisse müssen lückenlos und von einem fehlerfrei lauffähigen Programm mit dem Namen „*O\_master.do*“ erzeugt werden, welches alle im Projektverlauf relevanten Auswertungsprogramme enthalten muss. Dieses Programm muss zwingend mit dem Laden der vom 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 „*O\_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 „*O\_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 Nutzer/innen generierten Programmcodes (z.B. Stata ado-files) und deren jeweilige Version mit ein. Dies gilt auch für vom FDSZ bereitgestellte Software.
3. **Reproduzierbarkeit der Daten.** Alle zur Erzeugung der Berechnungsergebnisse genutzten Datensätze des FDSZ müssen zu Beginn der „*O\_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, dass 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 das FDSZ verlassen, sollen grundsätzlich dem Zwecke einer Publikation dienen. Daher sollen Gastforscher/innen 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 Gastforscher/innen 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 Gastforscher/innen 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 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 FDSZ benachrichtigt Gastforscher/innen sobald eine Zeilenanzahl von 2.000 überschritten wurde.
3. **Erhöhung der maximalen Zeilenanzahl für Output.** Gastforscher/innen 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 Gastforscher/innen 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 FDSZ die neue Zeilenanzahl für den Output fest. Sollte die Begründung nicht ausreichend sein, wird der Antrag auf Erhöhung der maximalen Zeilenanzahl für den Output vom FDSZ mit entsprechender Begründung abgelehnt. Eine einmalige Nachbesserung nach Ablehnung ist möglich.

4. **Maximale Zeilenanzahl für Output bei Projekten, an denen Forscher/innen der Deutschen Bundesbank beteiligt sind.** Nicht gezählt wird der Output, wenn Forscher/innen der Deutschen Bundesbank die Prüfung des Outputs selbst vornehmen. Diese Ausnahme gilt nicht für Projekte, bei denen das FDSZ die Outputprüfung durchführt bzw. durchführen muss.
5. **(IT-)ressourcenbewusste Programmierweise.** Gastforscher/innen 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

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

1. **Kontrolle aller herauszugebenden Berechnungsergebnisse.** Bevor die Gastforscher/innen Berechnungsergebnisse, außerhalb der gesicherten Umgebung verwenden können, muss zunächst durch das FDSZ kontrolliert werden, ob die Gastforscher/innen bei ihren Berechnungsergebnissen die Prinzipien zur Outputprüfung angewendet haben, und die Berechnungsergebnisse müssen vom FDSZ freigegeben werden. Gastforscher/innen müssen das FDSZ daher per E-Mail darüber informieren, ob und welche Berechnungsergebnisse geprüft werden sollen. Diese sollten 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 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 FDSZ nicht behoben und die Gastforscher/innen 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 weiter zu verarbeitenden Format vorliegen.

## 4 Prinzipien zur Publikationsprüfung

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

### Prinzip P.1: Kontrolle aller Publikationen durch das FDSZ

1. **Kontrolle aller Publikationen durch das FDSZ.** Forschungsergebnisse, die Berechnungsergebnisse beinhalten, dürfen erst veröffentlicht werden, wenn eine schriftliche Zustimmung der prüfenden Stelle 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.** Gastforscher/innen sind dafür verantwortlich jede angestrebte Veröffentlichung, die Forschungsergebnisse aus den Auswertungen während des Gastforscheraufenthaltes im FDSZ beinhaltet, dem FDSZ vorzulegen.
3. **Bezug auf zuvor freigegebener Berechnungsergebnisse / Reproduzierbarkeit der publizierten Ergebnisse.** Publikationen dürfen ausschließlich nur durch das FDSZ freigegebene Berechnungsergebnisse enthalten. Bei Einreichen einer Publikation zur Prüfung müssen Gastforscher/innen 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 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), 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, <Name des genutzten Einzeldatensatzes aus dem Standardangebot des FDSZ (gegebenenfalls einheitlich verwendete Abkürzung)>, <Verwendungszeitraum des genutzten Einzeldatensatzes>, 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 FDSZ stellt für fast alle der bereitgestellten Mikrodatsätze Datendokumentationen mit Metadaten bereit. Diese können auf der Website des FDSZs 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"><li>1. vor unbefugter und unrechtmäßiger Verarbeitung (Vertraulichkeit) und</li><li>2. vor unbeabsichtigtem Verlust, unbeabsichtigter Zerstörung oder unbeabsichtigter Schädigung (Integrität) bewahrt werden.</li></ol> 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 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 Erkennungskode, 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 85 % nicht übersteigen.
Dritte	Dritte sind alle Personen, die keine Gastforscher/innen oder Beschäftigte des 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 Identifikatoren aus dem Datenmaterial entfernt, der weitere Merkmalsumfang bleibt dagegen erhalten.
Forschungsergebnisse	Siehe Definition von <i>Berechnungsergebnisse</i> .
Gastforscherarbeitsplatz	Der Arbeitsplatz an dem Gastforscher/innen in der gesicherten Umgebung des FDSZs arbeiten können.
Gastforscher/innen	Alle Gastforscher/innen unterzeichnen einen Vertrag mit dem FDSZ und werden beim ersten Gastforscheraufenthalt verpflichtet. Im Rahmen ihres Forschungsprojektes haben Gastforscher/innen direkten Zugang zu vertraulichen Mikrodaten der Deutschen Bundesbank.
	...

Gast- forscheraufenthalt	Persönlicher Aufenthalt von Gastforscher/innen in den Räumlichkeiten des FDSZs. Gastforscher/innen müssen an einem Gastforscherarbeitsplatz in der gesicherten Umgebung arbeiten.
Gesicherte Um- gebung	In der gesicherten Umgebung können Gastforscher/innen an einem vom FDSZ 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 Identifika- tion	Identifikation von Berichtspflichtigen oder sonstige juristische oder natürliche Personen, Rechtssubjekten oder Niederlassungen durch Ableitung, sodass Einzelangaben bekannt werden.
Kontrollierte Daten- fernverarbeitung (KDFV)	Das FDSZ bietet für einige der bereitgestellten Mikrodatensätze eine kontrollierte Datenfernverarbeitung (KDFV) an. Mithilfe der KDFV können Forscher die Daten analysieren, ohne dabei physisch in den Räumen des FDSZ anwesend zu sein. Hierzu schicken sie ihre Auswertungsprogramme an das FDSZ.
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 fünf Merkmalsträger. Siehe Definition von <i>Fallzahl</i> .
Output	Siehe Definition von <i>Berechnungsergebnisse</i> .
Outputprüfung	Das FDSZ überprüft die Einhaltung der in diesem Dokument unter 3. Outputprüfung dargelegten Prinzipien und Richtlinien durch die Gastforscher/innen.
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 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 dem Bundesbank internen Laufwerk. Dieser ist mit der Projektnummer benannt und für Gastforscher/innen im Rahmen eines Gastforscheraufenthaltes zugänglich.
Prüfende Stelle	Das FDSZ oder einem anderen im Einzelfall für die Prüfung zuständigen Bereich der Bundesbank.

...

...



	...
Publikationsprüfung	Das FDSZ überprüft die Einhaltung der in diesem Dokument unter 4. Publikationsprüfung dargelegten Prinzipien und Richtlinien durch die Gastforscher/innen.
Re-Identifikation	Siehe Definition von <i>De-Anonymisierung</i> .
Scientific Use Files	Scientific Use Files (SUF) sind standardisierte Datensätze, die vom FDSZ für ausgewählte Statistiken erstellt werden. Durch die faktische Anonymisierung der Mikrodaten dürfen sie außerhalb der gesicherten Umgebung verwendet werden. SUF bieten im Vergleich zu einem Gastforscheraufenthalt ein geringeres Analysepotenzial, sind jedoch so konzipiert, dass sie sich für einen großen Teil der wissenschaftlichen Forschungsvorhaben eignen.
Veröffentlichung	<p>Als <b>Veröffentlichung</b> gilt unter anderem die Darstellung der Forschungsergebnisse.</p> <ul style="list-style-type: none"><li>– in Zeitschriftenartikeln,</li><li>– in Sammelbandbeiträgen,</li><li>– in Büchern,</li><li>– in Working Paper Series sowie Diskussionspapieren von Institutionen oder Unternehmen,</li><li>– in Dissertationen,</li><li>– bei jeglicher Online-Stellung, insbesondere auf Wissenschaftsplattformen wie SSRN.com und RePEc.org,</li><li>– in Konferenz-Webseiten oder persönlichen Webseiten wie etwa Google Pages etc.,</li><li>– in Mitarbeiterseiten von wissenschaftlichen Einrichtungen,</li><li>– in Einreichungen bei Konferenzen in Form von Arbeitspapieren,</li><li>– in allgemein zugänglichen Präsentationsfolien,</li><li>– in allgemein zugänglichen Bachelor- und Masterarbeiten.</li></ul> <p>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. <b>Nicht als Veröffentlichung</b> im Sinne des Absatzes 1 gilt die Darstellung der Forschungsergebnisse, sofern sie nicht online verfügbar sind,</p> <ul style="list-style-type: none"><li>– in mündlichen Vorträgen,</li><li>– in Präsentationsfolien,</li><li>– in Einreichungen bei Konferenzen (z.B. Kurztexpte, Abstracts),</li><li>– in Bachelor- und Masterarbeiten.</li></ul> <p>Für diese Darstellungen bedarf es keiner vorherigen schriftlichen Zustimmung des FDSZ oder eines anderen im Einzelfall hierfür zuständigen Bereichs der Bundesbank; die Geheimhaltungs- und Datenschutzerfordernungen gemäß diesen Bedingungen sind in diesen Fällen aber wie bei jeder Form der öffentlichen Darstellung einzuhalten.</p>
	...

Verpflichtung	Am Tag des ersten Gastforscheraufenthaltes werden Gastforscher/innen von Mitarbeiter/innen des FDSZ über die Verpflichtung nach § 1 Abs. 1 des Gesetzes über die förmliche Verpflichtung nichtbeamteter Personen (Verpflichtungsgesetz) vom 2. März 1974 und über die Verpflichtung auf das Datengeheimnis unterrichtet.
Zu schützende Einheit	Siehe Definition von <i>Merkmalsträger</i> .

---