

Dr. Johannes Rohde

hat Wirtschaftswissenschaften an der Leibniz Universität Hannover studiert und am dortigen Institut für Statistik im Jahr 2015 seine Promotion im Bereich Finanzmarktstatistik abgeschlossen. Seit 2017 ist er beim Landesbetrieb Information und Technik Nordrhein-Westfalen (IT.NRW) als Referent im Projekt Zensus 2021 tätig und bearbeitet dort alle methodischen Themen.

Christiane Seifert

studierte Wirtschaftswissenschaften an der Leibniz Universität Hannover und war im Anschluss mehrere Jahre in der Marktforschung tätig. Seit 2011 arbeitet sie bei IT.NRW und bearbeitete von 2015 bis 2017 im Forschungsdatenzentrum das Thema Geheimhaltung. Derzeit ist sie als Referentin für den Bereich Sondererhebungen verantwortlich.

Sarah Gießing

ist Diplom-Mathematikerin und leitet das Referat „Statistische Geheimhaltung; Mathematisch-statistische Methoden für Plausibilisierung und Imputation“ des Statistischen Bundesamtes. Sie ist Ko-Autorin internationaler und nationaler Standardwerke zur Statistischen Geheimhaltung und vertritt das Statistische Bundesamt im Exzellenzzentrum Statistische Geheimhaltung der Europäischen Union.

ENTSCHEIDUNGSKRITERIEN FÜR DIE AUSWAHL EINES GEHEIMHALTUNGSVERFAHRENS

Dr. Johannes Rohde, Christiane Seifert, Sarah Gießing

📌 **Schlüsselwörter:** Geheimhaltungsverfahren – Entscheidungskriterien – Verhaltenskodex für europäische Statistiken – Statistischer Verbund

ZUSAMMENFASSUNG

Um die Entscheidungsträger bei der Auswahl eines Geheimhaltungsverfahrens für eine Fachstatistik zu unterstützen, wurde ein Katalog von Entscheidungskriterien entwickelt. Dieser definiert zwanzig Kriterien, die ein Geheimhaltungsverfahren idealerweise erfüllen sollte, und prüft, inwiefern vier aktuell häufig diskutierte Verfahren diesen Kriterien genügen. In diesem Artikel wird der Kriterienkatalog vorgestellt, indem die dort beschriebenen geheimhaltungsrelevanten Kriterien diskutiert und deren gegenseitige Verflechtungen und Zielkonflikte herausgearbeitet werden. Alle Kriterien werden jeweils aus dem Verhaltenskodex für europäische Statistiken hergeleitet und lassen sich den Themenkomplexen Geheimhaltungsschutz, Datenqualität, Praktikabilität sowie Wirtschaftlichkeit zuordnen.

📌 **Keywords:** *statistical disclosure control methods – evaluation criteria – European Statistics Code of Practice – official statistics network*

ABSTRACT

A catalogue of evaluation criteria has been developed to support those responsible for selecting a statistical disclosure control (SDC) method for a given set of specialised statistics. It defines twenty criteria which an SDC method should ideally fulfil and assesses whether these criteria are satisfied by four frequently discussed methods. This article provides an introduction to the catalogue, presenting the criteria that are relevant to disclosure control and discussing their interdependencies and conflicting aims. All criteria have been derived from the European Statistics Code of Practice and grouped according to four main topics: protection against disclosure risks, data quality, feasibility, and cost effectiveness.

1

Einleitung

Der Schutz von Einzelangaben ist für die amtliche Statistik besonders in Hinblick auf die Veröffentlichung von Ergebnissen von sehr hoher Bedeutung. Die Gewährleistung der Geheimhaltung hat daher Eingang in den Verhaltenskodex für europäische Statistiken, den Code of Practice, gefunden¹ und ist national in § 16 Absatz 1 Bundesstatistikgesetz² geregelt.

Die konkrete Ausgestaltung der Geheimhaltung in einer Fachstatistik erfolgt durch das jeweils fachlich zuständige Gremium des Statistischen Verbunds³, die sogenannte Referentenbesprechung. Um die Entscheidungsträger bei der Auswahl eines Geheimhaltungsverfahrens unter Beachtung der methodischen Vorgaben zu unterstützen, wurde das Dokument „Entscheidungskriterien für die Auswahl eines Geheimhaltungsverfahrens“, der sogenannte Kriterienkatalog, entwickelt. Dieser bietet zunächst einen Überblick über geheimhaltungsrelevante Kriterien und prüft deren Erfüllung anschließend für vier ausgewählte Geheimhaltungsverfahren.

Der vorliegende Artikel stellt eine leicht gekürzte Fassung des ersten Abschnitts des Kriterienkatalogs dar. Darin erfolgt eine detaillierte Beschreibung von zwanzig Kriterien, die bei der Auswahl eines Geheimhaltungsverfahrens beachtet werden sollten. Hierbei werden verschiedene Aspekte praktischer, organisatorischer, technischer und rechtlicher Natur diskutiert und dabei die Anforderungen der Statistischen Ämter des Bundes und der Länder sowie der Nutzerinnen und Nutzer der Produkte der amtlichen Statistik berücksichtigt. Die Diskussion der Kriterien erfolgt stets vor dem Hintergrund, dass die Erfüllung eines jeden Kriteriums zwar wünschenswert ist, kein Geheimhaltungsverfahren dies jedoch gleichzeitig in aller Vollständigkeit gewährleisten

kann. Zum Teil bestehen dabei zwischen einzelnen Kriterien sowohl Synergieeffekte als auch Zielkonflikte unterschiedlichen Ausmaßes. Der Kriterienkatalog verfolgt nicht den Anspruch, fachstatistikspezifische Sonderfälle abdecken zu können. Aus diesem Grund wird keinerlei Priorisierung oder Kategorisierung nach Muss- und Soll-Kriterien vorgenommen, sondern stets auf allgemeiner Ebene argumentiert.

Im folgenden Hauptkapitel dieses Artikels werden geheimhaltungsrelevante Kriterien in vier Kategorien geordnet und detailliert beschrieben. Sofern möglich, erfolgt dabei jeweils eine Herleitung ihrer Relevanz aus dem Code of Practice. Zudem wird den Verflechtungen einzelner Kriterien durch entsprechende Verweise Rechnung getragen. Kapitel 3 zieht ein kurzes Fazit und bietet einen Ausblick auf das Gesamtdokument.

2

Kriterienkatalog

2.1 Gewährleistung der Statistischen Geheimhaltung (Schutz)

Der Schutz der Anonymität von Einzelangaben bildet einen Grundpfeiler im Zusammenspiel von Statistischen Ämtern und Auskunftgebenden. Der Code of Practice erhebt die Forderung nach der Gewährleistung der Geheimhaltung aller Angaben von Befragten und formuliert die Notwendigkeit einer gesetzlichen Regelung (Verhaltenskodex für europäische Statistiken, Grundsatz 5). In Deutschland sieht das Bundesstatistikgesetz grundsätzlich eine Pflicht zur Geheimhaltung von Einzelangaben vor (§ 16 Absatz 1 Bundesstatistikgesetz). Ausnahmen hiervon bestehen unter anderem dann, wenn die Einzelangaben von den Statistischen Ämtern mit den Einzelangaben anderer Befragter zusammengefasst veröffentlicht werden (§ 16 Absatz 1 Nr. 3 Bundesstatistikgesetz) oder den Befragten oder Betroffenen nicht mehr zuzuordnen sind (§ 16 Absatz 1 Nr. 4 Bundesstatistikgesetz). Deshalb sollte ein Geheimhaltungsverfahren gewährleisten, dass der Schutz der Einzelangaben der jeweiligen Fachstatistik unter den bestehenden Bedingungen (zum Beispiel hinsichtlich Auswertungsanforderungen oder Datenkonstellationen) gegeben ist. Bei der

1 Siehe dazu auch Klumpen/Schäfer, 2012.

2 Gesetz über die Statistik für Bundeszwecke (Bundesstatistikgesetz – BStatG) in der Fassung der Bekanntmachung vom 20. Oktober 2016 (BGBl. I Seite 2394), das zuletzt durch Artikel 10 Absatz 5 des Gesetzes vom 30. Oktober 2017 (BGBl. I Seite 3618) geändert worden ist.

3 „Statistischer Verbund“ bezeichnet den Verbund der Statistischen Ämter des Bundes und der Länder. Zum Statistischen Verbund gehören neben dem Statistischen Bundesamt die 14 Statistischen Landesämter. Bis auf das Statistische Amt für Hamburg und Schleswig-Holstein sowie das Amt für Statistik Berlin-Brandenburg verfügt jedes Bundesland über ein eigenes Statistisches Landesamt.

statistikspezifischen Operationalisierung der gesetzlichen Vorgaben sind vor Auswahl und Parametrisierung eines Geheimhaltungsverfahrens Anforderungen zur wirksamen Beschränkung bestehender Aufdeckungsrisiken zu formulieren. Die Kernaufgabe der Entwicklung eines Geheimhaltungskonzepts besteht darin, durch eine geeignete Verfahrens- und Parameterwahl die gesteckten Ziele bezüglich der Aufdeckungsrisiken einzuhalten.

2.1.1 Primäre Geheimhaltung

Die Notwendigkeit zur primären Geheimhaltung ergibt sich aus folgenden drei Risikokategorien (Gießing und andere, 2018, Abschnitt 2.2.1)⁴:

› Seltene oder einzigartige Merkmalskombinationen

Weisen Häufigkeitstabellen Felder mit sehr kleinen Fallzahlen auf, kann in bestimmten Fällen ein direkter Rückschluss auf einen Merkmalsträger möglich sein. Als geheimhaltungskritische Fallzahlen werden insbesondere Einzelfälle sowie Fallzahlen von Zwei betrachtet, da sich hier für die betroffenen Auskunftseinheiten gegebenenfalls die Möglichkeit bietet, mit dem Wissen um die eigene Merkmalsausprägung auf die Einzelangabe des jeweils anderen Befragten zu schließen. Weiterhin kann ein reales oder von den Betroffenen befürchtetes Aufdeckungsrisiko bestehen, wenn die Information aus dem Tabellenfeld aufgrund ihrer Einzigartigkeit mit Zusatzwissen verknüpft werden kann. Auch im Fall von Wertetabellen besteht bei schwachen Besetzungszahlen das Risiko einer Identifikation. Hier kommt hinzu, dass bestimmten Personen mit Zusatzwissen selbst bei Nicht-Veröffentlichung der kleinen Fallzahl bekannt sein kann, hinter welchen Werten einzelne beziehungsweise wenige statistische Einheiten stehen (Beispiel: Branchenkenner wissen, dass der Auslandsumsatz eines Wirtschaftszweigs von nur einem Unternehmen erwirtschaftet wird).

› Ausnutzung von Negativwissen zum Rückschluss auf Gruppenzugehörigkeit

Sind nur wenige Ausprägungen eines Merkmals besetzt, besteht in Häufigkeitstabellen grundsätzlich das Risiko einer Gruppenzuordnung der Befragten, das heißt das Zutreffen einer von wenigen möglichen Ausprägungen kann von Dritten sicher abgeleitet werden. Im Extremfall wird nur für eine einzige Merkmalsausprägung eine von Null verschiedene Häufigkeit beobachtet, während alle anderen Felder unbesetzt bleiben (Beispiel: Alle Einwohner einer Gemeinde sind katholischen Glaubens). Eine Gruppenzuordnung ist zudem möglich, falls die Differenz zwischen der Randsumme und einem Tabelleninnenfeld sehr gering ausfällt oder sogar lediglich Eins beträgt. So kann beispielsweise die einzige Person evangelischen Glaubens in einer Gemeinde mit dem Wissen um die eigene Ausprägung auf die Gruppenzugehörigkeit aller anderen Einwohner schließen, falls diese sämtlich der katholischen Kirche angehören. Außerdem ist die Zuordnung zu einer übergeordneten Gruppe möglich, wenn beispielsweise alle Gemeindemitglieder entweder der katholischen oder der evangelischen Kirche angehören und somit die Zugehörigkeit zu einer christlichen Kirche abgeleitet werden kann.

› Näherungsweise Rückschlüsse auf Einzelangaben

In Wertetabellen besteht das Risiko einer näherungsweisen Aufdeckung einer Einzelangabe, wenn sich die betreffende Merkmalssumme stark auf einen oder auf wenige Merkmalsträger konzentriert (Dominanzproblematik). Generell ist eine Abschätzung nach oben durch die Verwendung des Tabellenwertes als oberer Schranke mit gegebenenfalls hohem Schätzfehler immer möglich. Jedoch kann der Schätzfehler durch das Ausnutzen zusätzlichen Wissens um die Verteilung des Merkmals auf die Merkmalsträger unter Umständen beträchtlich verringert werden (beispielsweise bei Vorliegen eines Marktes mit allgemein bekannter hoher Konzentration).

⁴ Das „Handbuch zur Statistischen Geheimhaltung“ ist ein internes Dokument der Statistischen Ämter des Bundes und der Länder. Auf Anfrage kann das Dokument allgemein an andere nationale sowie internationale Statistikämter und andere Institutionen, die amtliche Statistiken produzieren (sogenannte ONAs, „other national statistical authorities“), sowie an andere Forschungsdatenzentren weitergegeben werden. Die Auflagen dabei sind, dass das Handbuch nur zur internen Verwendung bestimmt ist und nicht an Dritte weitergegeben werden darf. Über Anfragen von Interessenten außerhalb dieses Nutzerkreises wird im Einzelfall entschieden.

2.1.2 Sekundäre Geheimhaltung

Die Durchführung einer sekundären Geheimhaltung stellt die Aufrechterhaltung des primären Geheimhaltungsschutzes sicher. Grundsätzlich besteht die Möglichkeit, dass auf den primär geheim gehaltenen Wert mithilfe der Tabellenrandsummen durch einfache Dif-

ferenzbildung oder durch die Einbeziehung anderer, in logischer Beziehung zueinander stehender Tabellen zurückgeschlossen werden kann. Auf Basis der errechneten Differenzen oder der durch Tabellenvergleich abgeleiteten Unterschiede können somit Informationen über primär geschützte Einzelangaben gewonnen werden. Dies soll die sekundäre Geheimhaltung durch die Sperrung zusätzlicher Tabellenfelder verhindern.

Das Handbuch zur Statistischen Geheimhaltung präzisiert in Abschnitt 2.3.1 drei typische Enthüllungsszenarien durch Differenzbildung: Neben geografischen Differenzbildungsrisiken, die durch Abgleich verschiedener Tabellen mit identischem geografischen Bezug entstehen, besteht ein sekundäres Geheimhaltungsrisiko, wenn Tabelleninformationen separat für eine Auswahlgesamtheit und für eine in dieser enthaltenen Teilgesamtheit bereitgestellt werden, da so auf die Differenz beider Populationen zurückgeschlossen werden kann. Diese Tabelleninformationen können sowohl durch die Veröffentlichung einer hierarchisch gegliederten Tabelle (beispielsweise durch die Ausweisung von Statistiken mit identischer Bezugsgröße sowohl auf Landes- als auch auf Gemeindeebene) als auch durch die Veröffentlichung verschiedener, sich überlappender Tabellen (beispielsweise durch Angabe der Beschäftigtenzahlen nach Wirtschaftszweigen für alle Unternehmen einer Region und für alle Unternehmen ohne Auslandsumsätze derselben Region) bereitgestellt worden sein. Gemeinsam verwendet können Informationen aus verschiedenen Tabellen somit die Aufdeckung von Einzelangaben einer primär geheim gehaltenen Information durch Differenzbildung ermöglichen. Die Gesamtschutzwirkung eines Geheimhaltungskonzepts bestimmt sich somit durch die gleichzeitige Einhaltung des primären und sekundären Geheimhaltungsschutzes.

2.1.3 Einheitliche Geheimhaltung

Um den Schutz vor der Aufdeckung von Einzelangaben gewährleisten zu können, sollten alle Maßnahmen zur statistischen Geheimhaltung stets für das gesamte Veröffentlichungs- und Auswertungsprogramm (zum Beispiel Fachserien, Regionaldatenbank, GENESIS und so weiter) einheitlich geplant und über alle Anwendungsbereiche des Statistischen Verbunds hinweg konsistent umgesetzt werden (koordinierte Geheimhaltung). Eine getrennte Geheimhaltung verschiedener Veröffentlichungen mit identischen Bezugsgrößen genügt nicht,

da sonst beispielsweise in zwei verschiedenen Auswertungsprodukten zwar für sich allein genommen sichere, jedoch zusammen verwendet rückrechenbare Ergebnisse bereitgestellt werden könnten. Dieser Aspekt ist eng mit Kriterium 2.1.2 verknüpft, da sich sekundäre Geheimhaltungsrisiken als direkte Folge einer uneinheitlichen Geheimhaltung ergeben können.

Um eine Identifikation von Einzelangaben wirksam verhindern zu können, ist in allen Bereichen einer Fachstatistik somit die lückenlose Anwendung eines einheitlichen Geheimhaltungsverfahrens (einschließlich einer einheitlichen Parametrisierung) sowie die zeitliche Abstimmung der Geheimhaltung notwendig. Die Anwendung unterschiedlicher Verfahren zur Geheimhaltung birgt gegebenenfalls die Gefahr, dass sich die Geheimhaltungsmaßnahmen gegenseitig aufheben.

2.1.4 Geheimhaltung bei Nutzung von Einzeldaten durch Dritte

Eine zusätzliche Herausforderung an den unter Kriterium 2.1.3 formulierten einheitlichen Geheimhaltungsprozess entsteht bei der Übermittlung von Einzeldaten an Dritte oder falls diese Dritten selbst über die betreffenden Daten verfügen. Das Bundesstatistikgesetz benennt hier ausdrücklich die Wissenschaft (Kriterium 2.3.6) sowie zwei weitere mögliche anspruchsberechtigte Stellen und Einrichtungen der Datenübermittlung und regelt gleichzeitig die jeweilige Verwendung der Daten durch die externen Stellen.

An oberste Bundes- und Landesbehörden dürfen, soweit die spezialgesetzlichen Voraussetzungen vorliegen, Tabellen mit statistischen Ergebnissen übermittelt werden, auch wenn Tabellenfelder nur einen einzigen Fall ausweisen. Diese dürfen allerdings ausschließlich zur Verwendung gegenüber gesetzgebenden Körperschaften und für Planungszwecke (§ 16 Absatz 4 Bundesstatistikgesetz) genutzt werden, nicht aber zur Veröffentlichung. Für kommunale beziehungsweise andere in entsprechenden Rechtsvorschriften genannte Stellen, die Einzelangaben für statistische Zwecke erhalten, besteht im Falle einer Veröffentlichung eigener Auswertungen (zum Beispiel Zensus, statistisches Unternehmensregister, gegebenenfalls auch mit zusätzlichen, selbst erhobenen Merkmalen) hingegen die gleiche Verpflichtung zur Geheimhaltung wie für die amtliche Statistik (§ 16 Absatz 10 Bundesstatistikgesetz).

Wissenschaftliche Einrichtungen erhalten für Forschungszwecke einen Zugang zu Mikrodaten, das heißt zur Erstellung und Veröffentlichung von statistischen Ergebnissen (§ 16 Absatz 6 Bundesstatistikgesetz). Für diesen Nutzerkreis erfolgt die Datennutzung über die Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder (FDZ). Geheimhaltungsrelevante Aspekte aus Sicht der FDZ werden detailliert in Kriterium 2.3.6 benannt.

Um die Wirksamkeit von Geheimhaltungsmaßnahmen über Tabellen und Veröffentlichungen hinweg zu gewährleisten, ist es in den betroffenen Statistiken grundsätzlich wünschenswert, dass die Durchführung des Geheimhaltungsprozesses nicht nur im Statistischen Verbund, sondern auch bei den externen Stellen nach einheitlichen Maßstäben des Geheimhaltungsschutzes erfolgt (Kriterium 2.1.3). Somit sollte darauf hingearbeitet werden, dass externe Stellen in ihren Veröffentlichungen möglichst das identische Verfahren (einschließlich identischer Parameter) zur Geheimhaltung verwenden wie der Statistische Verbund. Alternativ sollten Vorgehensweisen vereinbart werden, die das Aufbrechen der Geheimhaltung der amtlichen Statistik vermeiden. Für betroffene Statistiken empfiehlt es sich, bei der fachlichen Festlegung des Geheimhaltungsverfahrens durch die Statistischen Ämter frühzeitig Kontakt mit der externen Stelle aufzunehmen und bei Bedarf entsprechende Beratung, möglicherweise sogar technische Unterstützung zur Verfügung zu stellen. Die Berücksichtigung dieses Aspekts erhöht bei den von den Ausnahmeregeln betroffenen Statistiken gegebenenfalls den innerhalb der Kriterien 2.4.1 bis 2.4.3 formulierten Implementierungs- und Umsetzungsaufwand.

2.2 Informationsverluste (Qualität)

Der Verhaltenskodex für europäische Statistiken formuliert wesentliche Anforderungen an die Qualität von statistischen Produkten (insbesondere in den Grundsätzen 11 bis 14). Da im Zuge der Sicherstellung der statistischen Geheimhaltung Einzelangaben geschützt werden müssen, ergibt sich unabhängig vom angewendeten Geheimhaltungsverfahren stets ein bestimmter Grad an Informationsverlust und damit auch ein Qualitätsverlust einer Statistik. Die Erhöhung des Schutzlevels sowie die damit verbundene Reduzierung des Aufdeckungsrisikos stehen somit in einem Zielkonflikt

mit der Sicherstellung einer möglichst hohen Qualität des statistischen Produkts.

2.2.1 Relevanz

Grundsatz 11 des Code of Practice fordert eine am Bedarf der Nutzerinnen und Nutzer orientierte Bereitstellung von Statistiken. Die Anwendung eines Geheimhaltungsverfahrens sollte somit gewährleisten, dass diese Interessen trotz des notwendigen Schutzes der Einzelangaben so weit wie möglich bedient werden können.

Grundsätzlich ist es die Aufgabe der Statistik, Massenerscheinungen darzustellen und nicht Einzelfälle abzubilden. Ein Geheimhaltungsverfahren sollte jedoch nicht der Auswertung auf einer für Nutzerinnen und Nutzer interessanten Gliederungstiefe, beispielsweise einer Veröffentlichung auf Gemeindeebene, im Wege stehen und damit zu einem unmittelbaren Verlust an Relevanz führen. Insbesondere bei kleinräumigen Analysen kann es zu sehr kleinen Fallzahlen und damit zu relativ großen Informationsverlusten aufgrund der Geheimhaltung kommen. Auch bei solch geringen Besetzungszahlen sollte das Geheimhaltungsverfahren einerseits anwendbar bleiben, gleichzeitig aber dem Erfordernis einer hinreichenden Ergebnisqualität Rechnung tragen. Abhängig vom verwendeten Geheimhaltungsverfahren kann die Relevanz zudem beeinträchtigt sein, wenn im Rahmen der Bund-Länder-Koordination des Geheimhaltungsprozesses die Geheimhaltung zu Lasten der oberen Ebenen erfolgen würde (siehe auch Kriterium 2.2.3). Neben einer ausreichenden Gliederungstiefe fällt unter das Kriterium der Relevanz auch die Verwertbarkeit der Ergebnisse: Werden die Parameter eines Geheimhaltungsverfahrens so gewählt, dass eine große Zahl an Werten gesperrt werden muss beziehungsweise die Daten sehr stark verändert werden, können die daraus resultierenden Ergebnisse für die Nutzenden ebenfalls an Relevanz verlieren (Kriterium 2.2.2).

Das Relevanzkriterium ist zudem eng mit dem Kriterium der Komplexität und Transparenz verzahnt (Kriterium 2.3.1), da ein intransparenter Geheimhaltungsprozess für die Nutzerinnen und Nutzer zu einer verminderten Relevanz des statistischen Produkts führen kann.

2.2.2 Genauigkeit

Eine amtliche Statistik soll die realen Verhältnisse genau und zuverlässig widerspiegeln (Verhaltenskodex für europäische Statistiken, Grundsatz 12). Ein Geheimhaltungsverfahren sollte die Genauigkeit der Ergebnisse daher nicht oder nur geringfügig beeinträchtigen. Aus diesem Grund sollte bei der Ausarbeitung der Verfahrensdetails beziehungsweise im Rahmen einer Parametrisierung das Ausmaß des Genauigkeitsverlusts verschiedener Verfahrensvarianten systematisch anhand geeigneter Kennzahlen bewertet werden⁵. Eine möglichst originalgetreue Abbildung des ursprünglichen Datensatzes sollte gewährleistet werden, um das Ziehen falscher Schlussfolgerungen aus statistischen Ergebnissen zu verhindern. Diese Gefahr besteht insbesondere dann, wenn eine Geheimhaltungsmaßnahme zu einer starken Veränderung in der Verteilung eines Einzelmerkmals führt oder den Zusammenhang zweier oder mehrerer Merkmale stark verzerrt. Je nach Verfahren können größere Veränderungen zufällig oder sogar in Form systematischer Verzerrungen auftreten, das heißt dem Verfahren inhärent sein. Letzteres Problem kann zum Beispiel eintreten, wenn Fallzahlen von 1 und 2 in einem Datensatz überrepräsentiert sind und deterministisch abgerundet werden. Ein nutzerseitiges Aufsummieren dieser Fallzahlen führt dann zu einer negativen Verzerrung der gebildeten Summe.

Kommt es bei einzelnen Ergebnissen zu einer starken Beeinträchtigung der Genauigkeit, sollte diese Qualitätseinschränkung für den Endnutzer klar gekennzeichnet sein (Kriterium 2.3.1). Abhängig vom konkreten Nutzerbedarf kann ein Genauigkeitsverlust auch eine verminderte Relevanz für die Nutzerin oder den Nutzer zur Folge haben (Kriterium 2.2.1).

Ein spezieller Sachverhalt besteht, wenn der Gesetzgeber unmittelbare Rechtsfolgen an das statistische Ergebnis knüpft (zum Beispiel im Falle von Einwohnerzahlen). Auch wenn im Einsatz von datenverändernden Verfahren⁶ grundsätzlich ein zulässiges Mittel zur Dar-

stellung von Ergebnissen der Bundesstatistik besteht, dürfen diese spezifischen statistischen Ergebnisse durch ein Geheimhaltungsverfahren nicht verändert werden.¹⁷ Bei der Wahl eines datenverändernden Verfahrens für eine Statistik muss in diesem Fall die Option bestehen, einzelne Werte, die nicht der primären Geheimhaltung unterliegen, unverändert zu veröffentlichen, ohne dass es dadurch zu inakzeptablen Aufdeckungsrisiken bei anderen statistischen Ergebnissen (zum Beispiel Einwohner nach Staatsangehörigkeit) kommt.

2.2.3 Aktualität

Grundsatz 13 des Code of Practice erhebt den Anspruch der Aktualität einer Statistik und fordert deren termingerechte Veröffentlichung. Aktualitätsverluste aufgrund von Geheimhaltungsmaßnahmen sollten deshalb im Zuge einer vorausschauenden Zeitplanung so weit wie möglich vermieden werden. Ein Geheimhaltungsverfahren kann dieser Forderung zum einen Rechnung tragen, indem dessen Durchführung in den Statistischen Ämtern zwar koordiniert und einheitlich, jedoch zeitlich unabhängig voneinander stattfinden kann. Dies ist insbesondere von Bedeutung, wenn die Erstveröffentlichung für eine Statistik durch die Länder erfolgt und die Veröffentlichung unabhängig vom Fortschritt des Geheimhaltungsprozesses anderer Länder erfolgen kann. Voraussetzung hierfür ist die Wahl eines entsprechenden Verfahrens sowie gegebenenfalls notwendiger Parameter. Für die bei bestimmten Verfahren zur Vermeidung von Aufdeckungsrisiken notwendige und sinnvollerweise zentral durchgeführte Geheimhaltung müssen hingegen erst die Daten aller Länder vorliegen, um den Geheimhaltungsprozess (vor allem hinsichtlich der Sekundärgeheimhaltung) abschließen zu können (Handbuch zur Statistischen Geheimhaltung, Abschnitt 2.4.5). Dies bedeutet für einzelne Länder mitunter einen Aktualitätsverlust. Die Möglichkeit einer automatisierten Durchführung des Geheimhaltungsverfahrens führt in der Regel zu deutlichen Zeitvorteilen gegenüber einer manuellen Bearbeitung und trägt somit zur Aktualität der Statistik und zur Einhaltung gesetzter Veröffentlichungstermine bei.

5 Eine Liste möglicher Kennzahlen zur Bewertung des Informationsverlusts von Geheimhaltungsverfahren sowie Beispielrechnungen zu verschiedenen Verfahren und Parametrisierungen enthält das Gesamtdokument „Entscheidungskriterien für die Auswahl eines Geheimhaltungsverfahrens“ (Rohde und andere, 2018).

6 Datenverändernde Verfahren (wie die stochastische Überlagerung oder SAFE) unterscheiden sich von informationsreduzierenden Verfahren (wie der Zellsperre) dadurch, dass sie geheim zu haltende Informationen nicht unterdrücken, sondern durch veränderte Ergebnisse ersetzen (Gießing und andere, 2014; Kleber/Gießing, 2018).

7 Auf Beschluss des Bund-Länder-Arbeitskreises „Rechtsfragen der Statistik“ im Jahr 2016.

2.2.4 Vergleichbarkeit

Grundsatz 14 des Verhaltenskodex für europäische Statistiken betont die Notwendigkeit der Kohärenz und Vergleichbarkeit statistischer Produkte. Zueinander in Beziehung stehende Statistiken sollen möglichst konsistent sein und gemeinsam verwendet werden können (Grundsatz 14.1). Ein Geheimhaltungsverfahren sollte diesen Grundsatz nicht substanziell beeinträchtigen. Übermittelt eine Fachstatistik einer anderen Fachstatistik Werte zur Veröffentlichung (zum Beispiel Umweltschutzinvestitionen, die als Teilposten in den Wert für allgemeine Investitionen einfließen), so sollte gewährleistet sein, dass der originäre Geheimhaltungsschutz nicht aufgehoben wird.

Ein weiterer wichtiger Nutzenaspekt vieler Statistiken besteht in der intertemporalen Vergleichbarkeit der betreffenden Statistik (Grundsatz 14.2). Die Anwendung eines Geheimhaltungsverfahrens sollte diese Vergleichbarkeit über den Zeitverlauf hinweg aufrechterhalten. Soweit möglich sollte sie nicht dazu führen, dass Datenlücken entstehen, starke Verzerrungen von Verhältniswerten auftreten oder Trends einer Zeitreihe signifikant verändert oder sogar umgekehrt werden (zum Beispiel durch eine stark gegenläufige Wertveränderung zu zwei aufeinanderfolgenden Berichtszeitpunkten – Kriterium 2.2.2). Da jeder Wechsel des Geheimhaltungsverfahrens im Zeitablauf faktisch einen Methodenbruch darstellt, sind dessen Auswirkungen möglichst gering zu halten und Nutzerinnen und Nutzern, wie auch etwaige auftretende starke Verzerrungen von Trends, klar zu kommunizieren (Kriterien 2.2.1 und 2.3.1).

2.2.5 Additivität

Ebenfalls durch Grundsatz 14 des Code of Practice gedeckt ist die Forderung nach der Additivität einer Tabelle. Diese liegt vor, wenn sich die Innenfelder einer Tabelle zeilen- und spaltenweise zur jeweiligen ausgewiesenen Randsumme addieren. Additivität beschreibt das Vorliegen tabelleninterner Konsistenz und steht damit in direktem Bezug zu den Kriterien 2.2.4 und 2.2.6. Nicht-Additivität entsteht, wenn die Geheimhaltung von Tabelleninnenfeldern und Randfeldern unabhängig voneinander erfolgt und somit Differenzen zwischen den aggregierten Ergebnissen und der entsprechenden ausgewiesenen Randsumme auftreten können.¹⁸

¹⁸ Nicht-Additivität ist von reinen Rundungsdifferenzen abzugrenzen.

Nicht-additive Ergebnisse werden von Nutzerseite unter Umständen nicht ohne Weiteres akzeptiert, weshalb einer klaren und transparenten Nutzerkommunikation eine besondere Bedeutung zukommt (Kriterium 2.3.1). Dies ist auch deshalb von Relevanz, da sich das Ausmaß der Nicht-Additivität immer weiter erhöhen kann, wenn nicht-additive Daten von Nutzerinnen und Nutzern nachträglich verdichtet werden. Dem Problem der daraus resultierenden Genauigkeitsverluste könnte mit der Veröffentlichung von Ergebnissen in vielfältigen Verdichtungsstufen entgegengetreten werden. Im Idealfall ermöglichen Web-Zugriffe sogar die Verwendung von Daten mit einem maßgeschneiderten Verdichtungsgrad (Kriterium 2.3.5). Werden Parameter oder Regeln des Geheimhaltungsprozesses im Zuge einer transparenten Darlegung des Geheimhaltungsprozesses kommuniziert, ist hingegen ein potenzielles Aufdeckungsrisiko einer geheimzuhaltenden Einzelangabe beim Auftreten von Nicht-Additivität zu beachten.

2.2.6 Konsistenz

Neben der in Kriterium 2.2.4 formulierten Vergleichbarkeit stellt die auswertungsübergreifende Konsistenz von Statistiken ein weiteres Qualitätskriterium statistischer Produkte dar, das sich aus Grundsatz 14.1 des Code of Practice ableiten lässt und auch nach der Anwendung eines Geheimhaltungsverfahrens erfüllt sein sollte. Logisch identische Angaben sollten über verschiedene Auswertungen hinweg aufrechterhalten und immer mit demselben Wert ausgewiesen werden. Eng mit diesem Aspekt verwoben ist die Vermeidung fachlicher Paradoxien in einer Statistik. Weisen Analysen logische Zusammenhänge auf, sollten nach Anwendung des Geheimhaltungsverfahrens keine in sich widersprüchlichen Ergebnisse entstehen (zum Beispiel das Ausweisen einer von Null verschiedenen Bevölkerungszahl mit Hochschulabschlüssen in der Altersklasse von 0 bis 5 Jahren).

Für Nutzerinnen und Nutzer sind konsistente Ergebnisse wesentlich, um eine eindeutige, nicht angreifbare Interpretation zum Beispiel von Studienergebnissen zu gewährleisten. Fehlende Konsistenz führt (wie auch Nicht-Additivität) bei Nutzerinnen und Nutzern zu Irritationen und erfordert eine transparente und durchdachte Kommunikation (Kriterium 2.3.1).

2.3 Datennutzung (Praktikabilität)

Um den in Grundsatz 15 des Verhaltenskodex für europäische Statistiken formulierten Ansprüchen der Zugänglichkeit und Klarheit zu genügen, sollte die amtliche Statistik einen benutzerfreundlichen Zugang zu ihren statistischen Produkten bieten und den Entstehungsprozess einer Statistik transparent darstellen. Dies bildet das Fundament für eine breite öffentliche Akzeptanz der amtlichen Statistik. Ergänzend zu den in Kapitel 2.2 diskutierten verbindlichen Qualitätsstandards benennt Grundsatz 15 des Code of Practice konkret das Prinzip einer klaren und verständlichen Präsentation sowie eine für die Endnutzung praktikable Art der Veröffentlichung der angebotenen Statistiken.

2.3.1 Komplexität und Transparenz

Zugänglichkeit und Klarheit erfordern laut Grundsatz 15.1 des Code of Practice unter anderem eine Präsentation von Statistiken und entsprechenden Metadaten, die sowohl deren korrekte Interpretation als auch aussagekräftige Vergleiche erleichtert. Als Anforderungen an ein Geheimhaltungsverfahren ergeben sich daraus einige eng miteinander verknüpfte Aspekte.

Zunächst sind Nutzerinnen und Nutzer darauf angewiesen, dass der Geheimhaltungsprozess möglichst transparent dargelegt werden kann. Sie benötigen klar nachvollziehbare Dokumentationen, in denen auch die essenziellen Vor- und Nachteile des eingesetzten Verfahrens benannt werden (Grundsatz 15.5 des Code of Practice). Diese sollten so abgefasst sein, dass der Geheimhaltungsschutz mithilfe der enthaltenen Informationen nicht rückgängig gemacht werden kann.

Um eine korrekte Interpretation der Ergebnisse zu ermöglichen, sollte eine verständliche Erläuterung der möglichen Einflüsse des Geheimhaltungsverfahrens auf die Datenqualität gemeinsam mit der Auswertung angeboten werden (Grundsatz 15.6 des Code of Practice). Zusätzlich zur erhebungsinhärenten Schätzungsgenauigkeit sollten die durch das Geheimhaltungsverfahren verursachten Einschränkungen in der Genauigkeit (insbesondere bei datenverändernden Verfahren) in jedem Fall kommuniziert und gegebenenfalls gekennzeichnet werden, damit Nutzerinnen und Nutzer entscheiden können, ob die Genauigkeit für ihre Zwecke ausreichend ist.

Inwieweit der Geheimhaltungsprozess verständlich und transparent dargelegt werden kann, hängt maßgeblich vom Grad der Komplexität des Verfahrens ab. Methodisch besonders komplexe Verfahren sind Nutzerinnen und Nutzern, aber auch innerhalb der Statistischen Ämter schwerer zu kommunizieren als sehr einfache oder bereits bekannte Verfahren. Infolgedessen steigt das Risiko, dass die zum Geheimhaltungsverfahren bereitgestellten Informationen nicht nachvollzogen werden können und dadurch gegebenenfalls sogar die Qualität des statistischen Produkts infrage gestellt wird. Die Akzeptanz des Geheimhaltungsverfahrens hängt somit unter anderem davon ab, inwieweit der in jedem Fall auftretende Informationsverlust für Nutzerinnen und Nutzer offensichtlich ist und ob der Sinn und Zweck des Informationsverlusts nachvollziehbar erläutert werden kann (zum Beispiel Inkaufnahme fehlender Additivität, um die Varianz der Veränderungen der Randsummen zu reduzieren). Darüber hinaus ist die transparente Kennzeichnung von Ergebnissen wichtig, deren Aussagekraft durch das Geheimhaltungsverfahren eingeschränkt oder nicht gegeben ist (Kriterium 2.2.2). Die Schaffung von Transparenz und einfacher Kommunizierbarkeit muss dabei stets unter der Maßgabe der Sicherstellung des Geheimhaltungsschutzes erfolgen (Abschnitt 2.1). So sollte berücksichtigt werden, dass beispielsweise die Kenntnis um die exakten Werte einzelner Parameter das Aufdeckungsrisiko erhöhen kann.

2.3.2 Replizierbarkeit

In Bezug auf die Arbeit mit empirischen Daten bedeutet Replizierbarkeit, dass eine Untersuchung bei identischer Datenbasis und unter Verwendung der gleichen Methodik wiederholbar ist und zu einem identischen Ergebnis führt. Die Reproduzierbarkeit von Ergebnissen ist für die Glaubwürdigkeit wissenschaftlicher Arbeiten von hoher Bedeutung, damit Analysen und Schlussfolgerungen von der wissenschaftlichen Gemeinschaft überprüft werden können (Kriterium 2.3.6). Für Veröffentlichungen der Statistischen Ämter stellt die Replizierbarkeit von Ergebnissen ebenfalls eine wichtige Eigenschaft dar, da das Vertrauen in amtliche Ergebnisse sinkt, wenn diese nicht von der unabhängigen Wissenschaft nachvollzogen werden können. Ein Geheimhaltungsverfahren sollte deshalb gewährleisten, dass auf einer identischen Datenbasis und auf gleicher Methodik beruhende Analysen weiterhin zu einem identischen Ergebnis führen.

2.3.3 Anwendungsbreite des Verfahrens

Je nach Statistik sind Fallzahlen und/oder Werte einer Geheimhaltungsprüfung zu unterziehen. Bei der Auswahl eines Geheimhaltungsverfahrens ist zunächst darauf zu achten, dass sich das Verfahren für die jeweilige Statistik eignet.

Neben der Geheimhaltung von Tabellen sollte sich ein Geheimhaltungsverfahren zur Anwendung auf möglichst viele weitere Formen statistischer Auswertungen eignen, wobei hier der Nutzerbedarf zu beachten ist (insbesondere bei FDZ-Nutzung, siehe Kriterium 2.3.6). Zu berücksichtigen ist außerdem, dass bei der Datennutzung eventuell neue Merkmale und Merkmalskategorien (zum Beispiel im Rahmen eines Web-Zugriffs auf Daten oder bei der Nutzung von Einzeldaten, siehe Kriterium 2.3.5) gebildet werden. Für die Anwendung des Geheimhaltungsverfahrens auf multivariate Analysen sollte sichergestellt werden, dass die Sinnhaftigkeit der errechneten Zusammenhänge, die Signifikanz der Effekte sowie die Interpretation der betreffenden Schätzwerte (zum Beispiel Regressionskoeffizienten) gemäß den Originaldaten erhalten bleiben. Bei der Erstellung von Zeitreihen sollte die intertemporale Vergleichbarkeit gewährleistet sein (Kriterium 2.2.4). Falls zudem eine Analyse auf geografischer Gitterzellenebene (§ 10 Absätze 2 und 3 Bundesstatistikgesetz) für die betreffende Statistik angeboten wird, sollte sich das Geheimhaltungsverfahren für tabellarische und kartografische Auswertungen gleichermaßen⁹ eignen, um eine einheitliche Geheimhaltung sicherzustellen (Kriterium 2.1.3). In allen beschriebenen Fällen ist gleichzeitig insbesondere auf die Erfüllung der in Abschnitt 2.2 formulierten Qualitätsstandards zu achten. Falls sich ein Geheimhaltungsverfahren nicht für alle genannten Auswertungsmethoden eignet, sollte zumindest eine Kombination mit einem anderen Geheimhaltungsverfahren möglich sein, wobei jedoch das Auftreten von zusätzlichen Aufdeckungsrisiken und auswertungsübergreifenden Inkonsistenzen zu vermeiden ist (Kriterium 2.2.4).

Weiterhin sollte ein Geheimhaltungsverfahren auch dann anwendbar bleiben, wenn amtliche Datensätze miteinander verknüpft oder mit externen Datensätzen

⁹ Das Expertengremium für die Analyse und Darstellung georeferenzierter Daten des Statistischen Verbunds hat 2017 beschlossen, dass „Geheimhaltung grundsätzlich das Gesamtveröffentlichungsprogramm berücksichtigen muss; tabellarische und kartografische Auswertungen müssen gemeinsam betrachtet werden“.

beziehungsweise zusätzlichen Merkmalen kombiniert werden, was insbesondere bei der Nutzung durch externe Dritte sowie in den Forschungsdatenzentren von Relevanz ist (Kriterien 2.1.4 und 2.3.6).

2.3.4 Flexibilität bei Veröffentlichungen

Als weiteren Indikator für die Zugänglichkeit eines statistischen Produkts benennt der Code of Practice in Grundsatz 15.3 die Bereitstellung von maßgeschneiderten Analysen. In der Regel lässt sich nicht im Voraus planen, welche Merkmalskombinationen einer Statistik für Veröffentlichungen oder Kundenanfragen benötigt werden. Ein Geheimhaltungsverfahren sollte daher ermöglichen, auch nicht zum abgestimmten Veröffentlichungsprogramm gehörende Analysen schnell und einfach einer Geheimhaltungsprüfung unterziehen zu können. Ansonsten fällt unter Umständen ein hoher (manueller) Geheimhaltungsaufwand an, da gleichzeitig auf eine einheitliche und konsistente Geheimhaltung mit den bereits veröffentlichten Ergebnissen zu achten ist (Kriterien 2.1.3 und 2.2.4). Das Erfordernis der Aufrechterhaltung des Geheimhaltungsschutzes innerhalb früherer Veröffentlichungen kann zu Einschränkungen in der Flexibilität hinsichtlich der Erstellung von Sonderanalysen oder bereits geplanter Auswertungen führen. In Konsequenz daraus kann es zu einem eingeschränkten Datenbeziehungsweise Merkmalsangebot kommen, welches wiederum die Frage nach der Relevanz des statistischen Produkts aufwerfen kann (Kriterium 2.2.1).

Ein hohes Maß an Flexibilität sollte auch hinsichtlich der Umsetzbarkeit des Verfahrens in den Statistischen Ämtern gegeben sein. Wie bereits in Kriterium 2.2.3 angesprochen, sollte ein Geheimhaltungsverfahren möglichst dezentral durchgeführt werden können, um Aktualitätsverluste so weit wie möglich zu vermeiden. Zudem kann eine dezentrale Durchführbarkeit der Geheimhaltung gegebenenfalls eine flexiblere Reaktion auf Kundenbedürfnisse ermöglichen.

2.3.5 Datenbanken und Web-Zugriff

Ein weiteres, mit dem Aspekt der Flexibilität (Kriterium 2.3.4) verbundenes Kriterium der praktikablen Datennutzung betrifft die Möglichkeit der Einbettung des Geheimhaltungsprozesses in die Struktur von Datenbanken, um einen flexiblen Nutzerzugriff anbieten zu können. Dieses Kriterium ist durch den Grundsatz 15.2

des Code of Practice motiviert, der eine Verbreitung der Statistiken mithilfe moderner Informations- und Kommunikationstechnologien fordert.

Ein Geheimhaltungsverfahren sollte in den Arbeitsablauf zur Befüllung von Datenbanksystemen integriert werden können. Das verwendete Verfahren sollte dabei effiziente Produktionsprozesse von Ergebnissen ermöglichen und die Aufdeckung von Einzeldaten für alle erstellbaren Auswertungen und über alle Aggregations Ebenen hinweg ausschließen.

Zudem sollte ein Geheimhaltungsverfahren die flexible Zusammenstellung von Auswertungen und den Abruf selbst definierter Datenquader für Nutzerinnen und Nutzer in Form von Web-Zugriffen auf die Datenbank, zum Beispiel über GENESIS oder die Zensusdatenbank, erlauben. Um den erforderlichen Geheimhaltungsschutz sicherzustellen, ist hierbei unter den Statistischen Ämtern eine enge Abstimmung bezüglich der eingesetzten Veröffentlichungsprogramme, Verfahren und Parameter erforderlich (Kriterium 2.1.3). Falls Web Reporting nicht nur den Abruf, sondern auch ein direktes Auslösen von Ergebnisberechnungen durch Nutzerinnen und Nutzer ermöglichen soll, muss der Geheimhaltungsmechanismus in den Berechnungsablauf integriert werden. Die verwendeten Werkzeuge müssen dann auf die Originaldaten zugreifen können. Hieraus ergibt sich gegebenenfalls ein höherer Implementierungsaufwand (Kriterium 2.4.1).

2.3.6 Wissenschaftliche Nutzung – Forschungsdatenzentren

Grundsatz 15.4 des Verhaltenskodex für europäische Statistiken fordert den Zugang zu Mikrodaten für Forschungszwecke. Dieses sogenannte Wissenschaftsprivileg wird für die deutsche amtliche Statistik in §16 Absatz 6 Bundesstatistikgesetz geregelt. Die wissenschaftliche Nutzung von Einzeldatensätzen erfolgt über die Forschungsdatenzentren.

Die Ergebnisse werden vor der Herausgabe an die Nutzerinnen und Nutzer in den Forschungsdatenzentren auf Geheimhaltung geprüft¹⁰, wobei die in der jeweiligen Fachstatistik angewandten Geheimhaltungsverfahren und -regeln analog angewandt werden (Kriterium 2.1.4).

¹⁰ Dies betrifft nur Ergebnisse, die über die sogenannte On-Site-Zugangswege erstellt werden.

Die zu den Fachstatistiken analoge Vorgehensweise in den Forschungsdatenzentren ist erforderlich, um nicht durch die Veröffentlichung unterschiedlich geheim gehaltener Ergebnisse Geheimhaltungsrisiken zu schaffen.

Zwar sind alle bisher beschriebenen Aspekte auch für die Geheimhaltung in den Forschungsdatenzentren relevant, einige allerdings besonders hervorzuheben: Da Nutzerinnen und Nutzer für Inhalt, Methode und Qualität der Ergebnisse selbst verantwortlich sind und diese die Grundsätze wissenschaftlichen Arbeitens einhalten müssen, sind in Bezug auf die Geheimhaltung eine hohe Transparenz der Geheimhaltungsmethodik, eine Qualitätskennzeichnung von Ergebnissen auf Basis datenverändernder Verfahren (Kriterium 2.3.1) sowie Konsistenz und Replizierbarkeit der Ergebnisse (Kriterien 2.2.6 und 2.3.2) unabdingbar.

Essenziell ist in den Forschungsdatenzentren die Anwendungsbreite von Geheimhaltungsverfahren (Kriterium 2.3.3), da die Wissenschaftsfreiheit ein selbstständiges und freies Arbeiten mit den Daten erfordert. So werden in den Forschungsdatenzentren beispielsweise statistische Analyseverfahren zur Erforschung von Zusammenhängen genutzt und verschiedenste statistische Kennzahlen erzeugt. Zudem werden Datensätze oftmals miteinander verknüpft. Geheimhaltungsverfahren sollten deshalb so konzipiert sein, dass dies weiterhin möglich ist. Dabei sollten sowohl Verknüpfungen von amtlichen Datensätzen untereinander als auch Verknüpfungen von amtlichen Daten mit Daten aus externen Quellen berücksichtigt werden.

Neben der On-Site-Nutzung in den Forschungsdatenzentren (Gastwissenschaftsarbeitsplatz, Kontrollierte Datenfernverarbeitung) werden der Wissenschaft auch faktisch anonyme Einzeldatensätze (Scientific-Use-Files) und absolut anonyme Einzeldatensätze (Public-Use-Files) zur Nutzung in der wissenschaftlichen Einrichtung zur Verfügung gestellt. Die Sicherstellung der Geheimhaltung erfolgt in diesem Fall nicht nach der Ergebniserstellung in den Forschungsdatenzentren, sondern a priori über die Anonymisierung der Einzeldatensätze durch die Forschungsdatenzentren. Für die ergebnisseitige Geheimhaltung und die Anonymisierung werden unter Umständen unterschiedliche Methoden eingesetzt (zum Beispiel Zellsperren zur Ergebnisgeheimhaltung, Vergrößerung zur Anonymisierung der Einzeldaten). Hierbei ist zu beachten, dass durch den Abgleich von

Veröffentlichungen kein Aufdeckungsrisiko entsteht, wenn die Ergebnisse über verschiedene Zugangswege erstellt und die zugrunde liegenden Daten damit unterschiedlich geheim gehalten beziehungsweise anonymisiert wurden.

Bei der Konzeption eines Geheimhaltungsverfahrens sollten die Bedarfe der wissenschaftlichen Nutzung berücksichtigt werden. Sollte ein neu entwickeltes Geheimhaltungsverfahren in den Forschungsdatenzentren nicht effizient einsetzbar sein, kann dies dazu führen, dass eine mit diesem Verfahren geheim gehaltene Statistik nicht (mehr) in den Forschungsdatenzentren angeboten werden kann und entsprechend für wissenschaftliche Auswertungen nicht zur Verfügung steht. Alternativ ist zu überlegen, ob in den Forschungsdatenzentren gegebenenfalls mit einem anderen Verfahren geheim gehalten werden kann, falls auch dieses effizient umsetzbar wäre und nicht zu einer Aufdeckung der geheim gehaltenen Daten führen würde. Damit würde allerdings vom bisherigen Grundsatz abgewichen, nach dem in den Forschungsdatenzentren immer analog zur Fachstatistik geheim gehalten wird (Kriterium 2.1.3). Vor diesem Hintergrund hat das den Referentenbesprechungen übergeordnete Gremium diese im Juni 2017 allgemein beauftragt, „die FDZ (...) bereits bei der Neukonzeption oder bei der Weiterentwicklung der jeweiligen Fachstatistik in die Diskussion zur Geheimhaltung einzubeziehen.“

2.4 Wirtschaftlichkeit

Ein effektiver Einsatz der zur Verfügung stehenden Ressourcen gewährleistet die Wirtschaftlichkeit statistischer Prozesse (Grundsatz 10 des Verhaltenskodex für europäische Statistiken). Alle Aufwände und Potenziale, die sich aus der Umsetzung eines Geheimhaltungsverfahrens in die Arbeitsprozesse einer Fachstatistik ergeben können, sind daher auch im Hinblick auf ihre Wirtschaftlichkeit zu prüfen.

2.4.1 Implementierungsaufwand

Die Umsetzung eines Geheimhaltungsverfahrens in den regulären Arbeitsbetrieb einer Fachstatistik erfordert im ersten Schritt Aufwand bei der methodischen Planung und Entwicklung eines statistikspezifischen Geheimhaltungskonzepts.

Während der Konzeptionsphase bedarf es vor allem genügend zeitlicher und personeller Ressourcen. Zunächst sollten dabei die Anforderungen der Fachstatistik (einschließlich rechtlicher Fragen) in Bezug auf die Geheimhaltung erarbeitet werden. Anhand dieser Anforderungen und des Kriterienkatalogs als Entscheidungshilfe kann anschließend ein passendes Geheimhaltungsverfahren ausgewählt werden. Berücksichtigt werden sollten dabei gegebenenfalls unterschiedliche Prioritäten von Bund und Ländern. Falls die Fachstatistik in den Forschungsdatenzentren bereitgestellt wird, sind diese zudem in den Planungsprozess einzubeziehen (Kriterium 2.3.6).

Der während der Konzeptionsphase anfallende Aufwand ist maßgeblich abhängig vom ausgewählten Geheimhaltungsverfahren: Lässt das Verfahren eine anschließende flexible Erstellung von Auswertungen entsprechend des Nutzerbedarfs nicht zu, müssen das Veröffentlichungsprogramm und der Detaillierungsgrad der Auswertungen bereits in der Konzeptionsphase festgelegt werden (Kriterien 2.3.4 und 2.3.5). Zudem kann je nach Verfahren ein zeitintensiver Planungsprozess für einen Geheimhaltungsmechanismus, der zwischen Bund und Ländern abgestimmt werden muss, notwendig sein.

Das entwickelte Konzept muss in einem weiteren Schritt geprüft und auch hinsichtlich der Parametrisierung des Verfahrens getestet werden. Hier entsteht unter Umständen zusätzlicher Aufwand, wenn der Evaluierungsprozess spezielle Expertise erfordert oder mit einer hohen Rechenzeit verbunden ist.

Nach der erfolgreichen Testphase können die einzelnen Schritte des Geheimhaltungsprozesses in geeigneter Weise in die Arbeitsprozesse der Fachstatistik implementiert werden, wofür bestehende Abläufe gegebenenfalls angepasst werden müssen. Der hierfür konkret anfallende Aufwand hängt von der eingesetzten Technik beziehungsweise der Verfügbarkeit von Modulen zur Umsetzung der Geheimhaltungsschritte in den vorgesehenen Auswertungsprogrammen ab und sollte bereits im Rahmen der Testphase abgeschätzt werden. Algorithmen und Funktionen, auf denen das Verfahren basiert beziehungsweise die zur Kennzeichnung von Ergebnissen verwendet werden, müssen in die Auswertungssoftware integriert werden beziehungsweise es sollten Codes oder Makros zur Umsetzung in der eingesetzten Softwarelösung zur Verfügung stehen. Falls entspre-

chende IT-Verfahren erst entwickelt und anschließend in den Arbeitsprozess der Fachstatistik implementiert werden müssen, entsteht kurz- und mittelfristig personeller, zeitlicher und gegebenenfalls monetärer Aufwand (Kriterium 2.4.3). Zu berücksichtigen sind außerdem die notwendige Einarbeitungszeit in die neue Methodik sowie gegebenenfalls entsprechende Schulungen für die betroffenen Mitarbeiterinnen und Mitarbeiter.

Wird die Fachstatistik in den Forschungsdatenzentren angeboten, sollte die Bereitstellung der zur Geheimhaltung verwendeten Software, Makros und Codes sowie die Unterstützung und Beratung der Forschungsdatenzentren bei der Implementierung gewährleistet sein, was zusätzlichen personellen und zeitlichen Aufwand sowohl in der Fachstatistik als auch in den Forschungsdatenzentren nach sich zieht.

2.4.2 Laufender Aufwand

Grundsätzlich sollte ein Geheimhaltungsverfahren so konzipiert sein, dass die geheimhaltungsspezifische Datenaufbereitung und die Freigabe der Ergebnisse zur Veröffentlichung kurzfristig nach der Ergebnisprüfung erfolgen kann (Kriterium 2.2.3 beziehungsweise Grundsätze 10.1 und 13 des Code of Practice). Für die Vorbereitung und die operative Durchführung der Geheimhaltung sollten daher laufend personelle und zeitliche Ressourcen eingeplant werden. Die Möglichkeit der automatischen Einbindung der Geheimhaltung in die IT-Prozesse ist deshalb zu berücksichtigen.

Einen für die Art und den Umfang des laufenden Aufwands entscheidenden Faktor stellt die Möglichkeit eines maschinellen Geheimhaltungsprozesses dar. Kann ein Verfahren nur manuell umgesetzt werden, sind der Umfang der zu prüfenden Tabellen sowie die Komplexität der erforderlichen Schritte hinsichtlich der zeitlichen Belastung von zentraler Bedeutung. Kann das Geheimhaltungsverfahren hingegen maschinell umgesetzt werden, stehen Automatisierungsgrad und anfallende Rechenzeit im Vordergrund. Die Rechenzeit hängt stark von der mathematischen Komplexität der eingesetzten Geheimhaltungsmethode ab. Bei sehr rechenintensiven Algorithmen fällt auch die Leistungsfähigkeit der zugrunde liegenden Hardware ins Gewicht. Können alle Schritte des Produktionsprozesses automatisiert ablaufen, verringert sich der Aufwand zur Sicherstellung der Geheimhaltung in den Fachabteilungen und den For-

schungsdatenzentren, was sich wiederum positiv auf die Wirtschaftlichkeit auswirkt. Zum laufenden Aufwand der Geheimhaltung gehört auch der eventuell anfallende Aufwand für eine Kennzeichnung hoher Abweichungen von den Originalwerten (Kriterium 2.2.2), die in diesem Fall möglichst automatisiert erfolgen sollte. Hierfür ist gegebenenfalls auch eine spezielle fachliche Beurteilung von Abweichungen bei Vergleichsindikatoren notwendig. Wie bereits für die Konzeptionsphase erörtert, ist auch für die Produktion von Auswertungen relevant, ob das Geheimhaltungsverfahren eine flexible Erstellung von Sonderauswertungen erlaubt (Kriterien 2.3.4 und 2.4.1). Ist dies nicht der Fall, können die Rechenzeit und auch die Zahl der Geheimhaltungsfälle durch die dann notwendige Bereitstellung von Auswertungen in vielfältigen Verdichtungsstufen stark ansteigen.

Personelle Ressourcen sollten außerdem für die Beratung und Kommunikation mit Nutzerinnen und Nutzern der statistischen Produkte einkalkuliert werden. Deren Umfang hängt von der Komplexität und vom Bekanntheitsgrad des Geheimhaltungsverfahrens ab. Insbesondere in den Forschungsdatenzentren besteht dazu gegebenenfalls auch fortlaufender Beratungsbedarf.

2.4.3 Potenzial für standardisierte IT-Lösungen zur Durchführung der Geheimhaltung

Die Standardisierung eines Geheimhaltungsprozesses kann zu einer Erhöhung der Wirtschaftlichkeit der Arbeitsabläufe einer Fachstatistik führen. Als Indikatoren für die Wirtschaftlichkeit statistischer Prozesse benennt der Verhaltenskodex für europäische Statistiken unter anderem die Ausschöpfung des Produktivitätspotenzials der Informations- und Kommunikationstechnologie (IT) bei der Datenverarbeitung (Grundsatz 10.2 des Code of Practice) sowie die Realisierung standardisierter Lösungen durch die statistischen Stellen (Grundsatz 10.4 des Code of Practice).

Um überhaupt Potenzial für eine methodische Standardisierung zu bieten, sollte ein Geheimhaltungsverfahren stets aus den gleichen Arbeitsschritten bestehen (Kriterium 2.4.2). In diesem Fall bietet sich die Möglichkeit, den Geheimhaltungsprozess mittels standardisierter Softwarelösungen zu unterstützen. Falls im Statistischen Verbund noch keine standardisierten Tools zur Verfügung stehen, sollten das Einsparpotenzial durch die Prozessstandardisierung und der Implementierungs-

aufwand, der durch die koordinierte Entwicklung entsprechender IT-Lösungen anfällt, gegeneinander abgewogen werden (Kriterium 2.4.1).

Zudem sollte geprüft werden, ob sich eine IT-Lösung direkt in vorhandene Auswertungsprogramme (wie zum Beispiel SAS) oder auch in eigene Produkte implementieren beziehungsweise mit ihnen kombinieren lässt (wie zum Beispiel GENESIS, länderspezifische Systeme). Für die Wirtschaftlichkeit ist außerdem von Relevanz, wie hoch der Automatisierungsgrad der IT-Lösung ist und in welchem Ausmaß noch manuell eingegriffen werden muss.

Im Rahmen einer Wirtschaftlichkeitsanalyse ist eine neue Software jedoch nicht nur in Hinblick auf ihre Funktionalität zu prüfen, sondern auch auf die Vereinbarkeit mit den technischen, organisatorischen, personellen und rechtlichen Rahmenbedingungen der deutschen amtlichen Statistik. Hierbei sind gegebenenfalls die Bewertung durch das Expertengremium für Geheimhaltung im Statistischen Verbund und die Erfahrungen von Teststatistiken zu beachten.

2.4.4 Nutzbarkeit des Potenzials der Informations- und Kommunikationstechnologie bei der Datenverbreitung

Der Code of Practice fordert in Grundsatz 10.2 zur Sicherstellung der Wirtschaftlichkeit die Ausschöpfung des Potenzials moderner Informations- und Kommunikationstechnologien nicht nur bei der Verarbeitung von Daten (Kriterium 2.4.3), sondern auch bei deren Verbreitung. Diese Technologien ermöglichen den Nutzerinnen und Nutzern flexible und in weiten Teilen selbst spezifizierbare Auswertungen des Datenangebots. Deshalb ist zu prüfen, ob die jeweilige Geheimhaltungsmethodik dies bei der Datenverbreitung gewährleisten kann. Ist diese Eigenschaft erfüllt, fällt in den Ämtern kein zusätzlicher personeller oder zeitlicher Aufwand für die separate Geheimhaltung von Sonderanalysen an. Voraussetzung für Wirtschaftlichkeitsgewinne ist, dass das Geheimhaltungsverfahren vollautomatisch umgesetzt ist. In Abgrenzung zu Kriterium 2.3.4 liegt der Fokus dieses Aspekts auf der Flexibilität aus Sicht der Statistischen Ämter (und nicht aus Nutzersicht).

Das Potenzial der Informations- und Kommunikationstechnologie kann in Abhängigkeit vom genutzten Geheimhaltungsverfahren einerseits und den für

Berechnungen in einem Vertriebssystem vorgehaltenen Daten andererseits unterschiedlich gut ausgeschöpft werden. Wir unterscheiden hier drei Szenarien mit unterschiedlich hohem Ausschöpfungspotenzial:

- › Zu einem optimalen Ausschöpfungspotenzial kommt es, wenn in einem Vertriebssystem Originaldaten vor Geheimhaltung hinterlegt werden können und Relevanz und Genauigkeit der Ergebnisse durch die Geheimhaltung nicht nennenswert beeinträchtigt werden. Ein solches System stellt jedoch hohe Anforderungen an die IT-Sicherheit und wurde im Statistischen Verbund bisher nicht implementiert. Werden im Vertriebssystem trotzdem (durch geeignete IT-Sicherheitsvorkehrungen abgesichert) Originaldaten unverändert eingelagert und abgerufene Ergebnisse dann auf dieser Datengrundlage berechnet, sind Relevanz und Genauigkeit der Ergebnisse bei Geheimhaltung mit einem post-tabularen Verfahren¹¹ jedoch in der Regel sichergestellt. Würden dagegen ausschließlich prä-tabular¹² geheim gehaltene Daten eingelagert, könnten die Anforderungen an die IT-Sicherheit reduziert werden. Es bestünde dabei aber die Gefahr, dass Relevanz und Genauigkeit eines Teils der Ergebnisse nach Geheimhaltung den fachlichen Ansprüchen nicht mehr vollumfänglich genügen. Auch im Fall der Nutzung von prä-tabular geheim gehaltenen Daten kann es deshalb sinnvoll sein, zusätzlich auch Originaldaten einzulagern, um eine automatische Kennzeichnung von Ergebnissen mit hohen Abweichungen zu ermöglichen. Falls eine derartige Kennzeichnung möglich wäre, fielen sowohl für den prä-tabularen als auch für den post-tabularen Ansatz kein zusätzlicher personeller oder zeitlicher Aufwand in den Ämtern an, um Sonderanalysen geheim zu halten beziehungsweise um große Abweichungen zu kennzeichnen. Im Hinblick auf Aspekte der IT-Sicherheit wären diese beiden Varianten als gleichwertig anzusehen.
- › Ohne hinterlegte Originaldaten kann ein reduziertes, aber noch immer hohes Ausschöpfungspotenzial erreicht werden. Die Auswertungsmöglichkeiten

11 Post-tabulare Geheimhaltungsverfahren (zum Beispiel Rundung oder stochastische Überlagerung) setzen erst bei den fertigen statistischen Ergebnissen (zum Beispiel in Tabellen ausgewiesene Häufigkeiten) an.

12 Prä-tabular datenverändernde Geheimhaltungsverfahren (zum Beispiel SAFE) werden direkt auf den Originaldatenbestand angewendet und modifizieren diesen so, dass für alle vorgesehenen Tabellen gleichzeitig die Anforderungen der Geheimhaltung erfüllt werden.

beschränken sich dann auf fest definierte Quader und daraus entweder dynamisch berechenbare oder in entsprechendem Umfang auf vielfältigen Aggregationsebenen hinterlegte, dynamisch abrufbare Ergebnisse. Die Relevanz und Genauigkeit der Ergebnisse kann in diesem Szenario gut gewährleistet werden. Die Flexibilität der Auswertung ist allerdings gegenüber einem System mit hinterlegten Originaldaten etwas eingeschränkt.

- › Ein demgegenüber beschränktes Ausschöpfungspotenzial bieten Systeme, die nur Zugriffe auf fest hinterlegte Quader (in relativ wenigen, grundsätzlich hierarchisch angeordneten Aggregatstufen) ermöglichen und insbesondere keine nutzerseitig auslösbaren dynamischen Berechnungen zulassen. Die Genauigkeit der Ergebnisse kann auch in diesem Szenario sichergestellt werden. Die Relevanz kann allerdings durch den begrenzten Umfang an möglichen Auswertungen weiter eingeschränkt sein. Die Anforderungen an die IT-Sicherheit sind für die Szenarien eines hohen und eines beschränkten Ausschöpfungspotenzials geringer als bei Nutzung eines optimalen Ausschöpfungspotenzials.

Bei der Diskussion zur Festlegung eines Geheimhaltungsverfahrens ist zu beachten, dass das Kriterium des Ausschöpfungspotenzials nicht nur in Zusammenhang mit der Relevanz und der Genauigkeit der Ergebnisse steht (Kriterien 2.2.1 und 2.2.2), sondern auch mit den übrigen der in Abschnitt 2.4 genannten Kriterien der Wirtschaftlichkeit eines Geheimhaltungsverfahrens verknüpft ist.

3

Fazit und Ausblick auf den Kriterienkatalog

Dieser Beitrag formuliert relevante Kriterien, die für die Auswahl eines Geheimhaltungsverfahrens in einer Fachstatistik berücksichtigt werden sollten, und ordnet diese vier Kategorien zu. Diese detaillierte Beschreibung der Kriterien stellt einen wesentlichen Teil des ersten Kapitels des Kriterienkatalogs dar. In diesem werden weiterhin vier aktuell häufig diskutierte Geheimhaltungsverfahren (Zellspernung, Deterministische Rundung,

SAFE und Stochastische Überlagerung) hinsichtlich der beschriebenen Kriterien beurteilt und miteinander verglichen. Das Gesamtdokument wurde im Februar 2018 durch das zuständige Lenkungsgremium des Statistischen Verbunds verabschiedet und soll Referentenbesprechungen künftig als verbindlicher Rahmen für die Entscheidungsfindung für ein Geheimhaltungsverfahren dienen.¹³ Ergänzend zum Kriterienkatalog bieten das Handbuch zur Statistischen Geheimhaltung beziehungsweise das im Rahmen der Europäischen Kooperation entstandene Handbuch „Statistical Disclosure Control“ (Hundepool und andere, 2012) eine ausführliche Darstellung von Geheimhaltungsmethoden. [UU](#)

¹³ Eine aktuelle Version findet sich im Intranet der Statistischen Ämter des Bundes und der Länder (Rohde und andere, 2018).

LITERATURVERZEICHNIS

Gießing, Sarah/Habla, Heike/ Höninger, Julia/Hoffmeister, Rita/Merz, Franz-Josef/Richter, Alexander/Scharnhorst, Sebastian/Schmidtke, Kerstin/Spies, Lydia/Tonte, Andreas/Uhrich, Stefanie. *Handbuch zur Statistischen Geheimhaltung. Stand 12. April 2018*. Internes Dokument des Statistischen Verbunds, verfügbar im Intranet der Statistischen Ämter des Bundes und der Länder.

Gießing, Sarah/Heinzl, Felix/Kleber, Birgit/Wilke, Achim. [Geheimhaltung beim Zensus 2011](#). In: *Wirtschaft und Statistik*. Ausgabe 11/2014, Seite 641 ff.

Hundepool, Anco/Domingo-Ferrer, Josep/Franconi, Luisa/Giessing, Sarah/Schulte Nordholt, Eric/Spicer, Keith/de Wolf, Peter-Paul. *Statistical Disclosure Control*. Wiley Series in survey methodology. Chichester 2012. Überarbeitete Fassung des ESSNET-SDC Handbook on Statistical Disclosure Control von 2010.

Kleber, Birgit/Gießing, Sarah. [Geheimhaltung beim Zensus 2021](#). In: *WISTA Wirtschaft und Statistik*. Ausgabe 3/2018. Kurznachrichten, Seite 6 ff., hier: Seite 7 f.

Klumpen, Dorothea/Schäfer, Dieter. [Der Verhaltenskodex für europäische Statistiken \(Code of Practice\) in überarbeiteter Fassung 2011](#). In: *Wirtschaft und Statistik*. Ausgabe 12/2012, Seite 1035 ff.

Rohde, Johannes/Seifert, Christiane/Gießing, Sarah/Setzer, Stefanie unter Mitarbeit von Breitenfeld, Jörg/Brings, Stefan/Höhne, Jörg/Höninger, Julia/Rothe, Patrick/Schedding-Kleis, Ulrike. *Entscheidungskriterien für die Auswahl eines Geheimhaltungsverfahrens. Version 1.0 vom 18.4.2018*. Internes Dokument des Statistischen Verbunds, verfügbar im Intranet der Statistischen Ämter des Bundes und der Länder.

Verhaltenskodex für europäische Statistiken für die nationalen und gemeinschaftlichen statistischen Stellen, angenommen vom Ausschuss für das Europäische Statistische System. 28. September 2011. [Zugriff am 9. Mai 2018]. Verfügbar unter: <http://ec.europa.eu/eurostat>

RECHTSGRUNDLAGEN

Gesetz über die Statistik für Bundeszwecke (Bundesstatistikgesetz – BStatG) in der Fassung der Bekanntmachung vom 20. Oktober 2016 (BGBl. I Seite 2394), das zuletzt durch Artikel 10 Absatz 5 des Gesetzes vom 30. Oktober 2017 (BGBl. I Seite 3618) geändert worden ist.

Herausgeber

Statistisches Bundesamt (Destatis), Wiesbaden

Schriftleitung

Dr. Sabine Bechtold

Redaktionsleitung: Juliane Gude

Redaktion: Ellen Römer

Ihr Kontakt zu uns

www.destatis.de/kontakt

Erscheinungsfolge

zweimonatlich, erschienen im Juni 2018

Das Archiv aller Ausgaben ab Januar 2001 finden Sie unter www.destatis.de/publikationen

Print

Einzelpreis: EUR 18,- (zzgl. Versand)

Jahresbezugspreis: EUR 108,- (zzgl. Versand)

Bestellnummer: 1010200-18003-1

ISSN 0043-6143

ISBN 978-3-8246-1070-9

Download (PDF)

Artikelnummer: 1010200-18003-4, ISSN 1619-2907

Vertriebspartner

IBRo Versandservice GmbH

Bereich Statistisches Bundesamt

Kastanienweg 1

D-18184 Roggentin

Telefon: +49 (0) 382 04 / 6 65 43

Telefax: +49 (0) 382 04 / 6 69 19

destatis@ibro.de

Papier: Metapaper Smooth, FSC-zertifiziert, klimaneutral, zu 61% aus regenerativen Energien

© Statistisches Bundesamt (Destatis), 2018

Vervielfältigung und Verbreitung, auch auszugsweise, mit Quellenangabe gestattet.