



Alexander Kuhnt
Referat Bevölkerungs- und
Haushaltsstatistiken

Die Geheimhaltung mittels Cell-Key-Methode

Auswirkungen auf die Ergebnisdarstellung der Bevölkerungsstatistiken

1. Einleitung

Bei den Bevölkerungsstatistiken des statistischen Verbundes stehen ab dem Berichtsjahr 2025 große Veränderungen an. Die Geheimhaltung der Ergebnisse wird in den meisten Bevölkerungsstatistiken auf ein neues Verfahren umgestellt, auf die so genannte Cell-Key-Methode. Zu den betroffenen Bevölkerungsstatistiken zählen die Statistik der Einbürgerungen, der Geburten, der Sterbefälle, der Eheschließungen sowie die Wanderungsstatistik (Zu- und Fortzüge). Die Statistik der Fortschreibung der Bevölkerung – und damit die vom Thüringer Landesamt für Statistik veröffentlichten Einwohnerzahlen der Gemeinden und Kreise Thüringens – ist aufgrund der Methodik¹⁾ jedoch nicht von dieser Umstellung betroffen, weshalb es hierzu keinerlei weiterer Ausführung bedarf.

Auch im Thüringer Landesamt für Statistik laufen die Vorbereitungen auf die Umstellungen seit Längerem. In diesem Aufsatz sollen die Veränderungen bei der Ergebnisdarstellung in den Bevölkerungsstatistiken infolge der Einführung der Cell-Key-Methode zur Geheimhaltung der Daten sowie die Vor- und Nachteile der Verwendung dieser Methode aufgezeigt werden. Anhand einer sehr vereinfachten Darstellung, wie die Cell-Key-Methode funktioniert, sollen die anstehenden Veränderungen aufgezeigt werden. Zuvor erfolgt jedoch die Betrachtung des Status Quo der Geheimhaltung in den Bevölkerungsstatistiken.

2. Warum werden Geheimhaltungsmaßnahmen in den Bevölkerungsstatistiken benötigt?

Bei den genannten Bevölkerungsstatistiken werden vorrangig Daten der Verwaltungsbehörden verarbeitet. So melden die Einwohnermeldeämter die Zu- und Fortzüge über die Gemeindegrenzen an die Statistik. Zudem werden Angaben zu Staatsangehörigkeitswechseln geliefert. Die Standesämter melden die Geburten, Sterbefälle und Eheschließungen.

Von den Ausländer- und Einbürgerungsbehörden erhält die amtliche Statistik die Daten zu den Einbürgerungen.

Schutz der persönlichen Daten ist eines der wichtigsten Anliegen der amtlichen Statistik

Auch bei den Bevölkerungsstatistiken ist der Schutz der persönlichen Daten, ebenso wie bei der amtlichen Statistik im Allgemeinen, eines der wichtigsten Anliegen. Die Personen, deren Verwaltungsdaten an die amtliche Statistik gesendet werden, müssen darauf vertrauen können, dass die Auswertungen und Veröffentlichungen dieser Daten keine Rückschlüsse auf die persönlichen Verhältnisse und Gegebenheiten des Betroffenen ermöglichen. Die amtliche Statistik ist daher per Gesetz verpflichtet²⁾, mit geeigneten Maßnahmen dafür Sorge zu tragen, dass die Interessen aller Beteiligten, Datennutzer wie Datengeber, gewahrt bleiben.

Das macht jedoch Auswertungen in tief strukturierter, sehr fein gegliederter Form sowie bei kleinen regionalen Einheiten oder bei einer kleinen Grundgesamtheit schwierig bis unmöglich. Dabei spielt es für die amtliche Statistik keine Rolle, ob die Information möglicherweise bereits woanders in einer öffentlich zugänglichen Quelle veröffentlicht wurde. So gab es in Thüringen beispielsweise in den vergangenen Jahren Mehrlingsgeburten mit jeweils 4 Kindern. Diese besonders freudigen Ereignisse wurden medial begleitet und in lokalen Nachrichten gezeigt sowie in der lokalen Presse abgedruckt, jeweils mit Namen und Bildern zu den Eltern und den jeweils 4 Kindern, teilweise mit Information zum Alter oder zur nationalen Herkunft der Eltern. Dies geschah vermutlich mit Zustimmung der Betroffenen.

1) Die Ergebnisse der Statistik der Bevölkerungsfortschreibung werden auf der Ebene von Kohorten errechnet. Die Zahlen sind nicht auf die Angaben von Einzelpersonen zurückzuführen.

2) siehe u. a. §16 Bundesstatistikgesetz

Die amtliche Statistik darf per Gesetz eine solche feingliedrige Auswertung nicht veröffentlichen, auch wenn es dem Dienstleistungs- und Informationsgedanken entspricht, dass man als statistische Behörde gern so genau und detailliert wie möglich auswerten und berichten möchte. Grundsätzlich müssen persönliche Angaben und Einzelangaben geschützt werden, auch bei den Bevölkerungsstatistiken.

3. Wie wurde die Geheimhaltung in den Bevölkerungsstatistiken bisher gehandhabt?

Um Einzelangaben in der Darstellung statistischer Ergebnisse zu vermeiden, gibt es mehrere recht einfache Verfahren, welche zum Teil in den Bevölkerungsstatistiken seit vielen Jahrzehnten angewendet werden. Darunter zählen Maßnahmen wie Vergröberung, Aggregation oder die Zellspernung.

Aggregation von Daten ist eine einfache Geheimhaltungsvariante

Um die Geheimhaltung in den Bevölkerungsstatistiken sicherzustellen, war der erste, einfachste und auch am häufigsten verwendete Schritt, die Daten auf ein gröberes Auswertungsniveau zu aggregieren. Statt dem Einzelalter wurde beispielsweise die Auswertung mit Altersgruppen durchgeführt. Statt einzelne Staatsangehörigkeiten auszuweisen wurde nur nach Nationalität (deutsch/nichtdeutsch) ausgewertet. Statt nach einzelnen Gemeinden wurden Gruppen gebildet oder die Auswertung auf die Ebene der Kreise gehoben. Auch das bewusste Weglassen bestimmter Spalten, beispielsweise bei der Auswertung von Zu- und Fortzügen nur den errechneten Saldo statt der Zu- und Fortzugszahlen auszuweisen, war in manchen Fällen sachdienlich. Hat das Aggregieren oder Weglassen nicht geholfen, kamen weitere Verfahren zur Anwendung, welche sich jedoch von Statistik zu Statistik unterschieden haben.

Bei der Fünfer-Rundung ist keine Aufrechnung möglich

In der Einbürgerungsstatistik und in den Ergebnissen zum Ausländerzentralregister wurden in den vergangenen Jahren alle Zahlen auf das nächste Vielfache von 5 gerundet. Dies hatte den Vorteil, dass

die Durchführung der Geheimhaltung relativ simpel und eine Rückrechnung auf die originalen Ergebnisse nicht möglich war. Die Daten konnten zudem sehr fein strukturiert mit vielen Merkmalsausprägungen ausgewertet werden. Nachteilig war, dass eine 5er-Rundung ein recht grobes Verfahren ist. Zudem ist keine Aufrechnung möglich, was bei kleinen Tabellen zu absurd anmutenden Ergebnisdarstellungen führen kann. Wie dem Beispiel aus Abbildung 1 zu entnehmen ist, kann es bei einer Tabelle mit 4 besetzten Feldern mit jeweils einem Wert von 3 oder 4 dazu kommen, dass infolge der Fünfer-Rundung eine nicht aufrechenbare Gleichung entsteht. Die Summe aus den einzelnen Wert $3+3+3+4$ ergibt 13. Werden alle Werte der Tabelle auf ein Vielfaches von 5 gerundet, ergibt sich folgendes Bild: $5+5+5+5=15$. Das ist infolge der Geheimhaltung eine korrekte Darstellung, aber für den Außenstehenden ist dies natürlich schwer zu erklären, wenn dieser die originalen Werte nicht kennt.

Abbildung 1: Beispiel einer Einbürgerungstabelle ohne und mit Geheimhaltung (5er-Rundung)

Einbürgerungen nach der bisherigen Staatsangehörigkeit (ohne Geheimhaltung)	
Land	Personen
Mexiko	3
Bolivien	3
Peru	3
Brasilien	4
Summe	13

Einbürgerungen nach der bisherigen Staatsangehörigkeit (mit Geheimhaltung - 5er Rundung)	
Land	Personen
Mexiko	5
Bolivien	5
Peru	5
Brasilien	5
Summe	15

Bei den Statistiken zu den Geburten, Sterbefällen, Eheschließungen und den Wanderungen wurde in den letzten Jahren das Verfahren der Zellspernung bei zu kleinen Werten angewendet. Es galt die Regel, dass ein Feld mindestens mit der Fallzahl 3 belegt sein musste, um einen Rückschluss auf die Angaben einer Einzelperson auszuschließen oder zumindest

deutlich zu erschweren. 3 Fälle sind die Mindestzahl, da auf diese Weise ein Rückschluss auf eine dritte Person nicht möglich ist, selbst wenn man selbst eine der betroffenen 3 Personen ist. Werte kleiner 3 wurden grundsätzlich gesperrt. Um die Ausrechenbarkeit der Felder zu verhindern, wurden zudem mindestens ein weiteres Feld in derselben Spalte bzw. Zeile gesperrt (sekundäre Sperrung).

Abbildung 2: Beispiel einer Tabelle der Geburtenstatistik ohne und mit Geheimhaltung (Zellsperrung/Aggregation)

Geburten nach Einzel- und Mehrlingsgeburten (ohne Geheimhaltung)	
Merkmal	Personen
Einzelgeburt	13 993
Zwillinge	421
Drillinge	3
Vierlinge	1
Summe	14 418

Geburten nach Einzel- und Mehrlingsgeburten (mit Geheimhaltung - Zellsperrung)	
Merkmal	Personen
Einzelgeburt	13 993
Zwillinge	421
Drillinge	.
Vierlinge	.
Summe	14 418

Geburten nach Einzel- und Mehrlingsgeburten (mit Geheimhaltung - Aggregation)	
Merkmal	Personen
Einzelgeburt	13 993
Mehrlingsgeburten	425
Summe	14 418

Bei kleinen Fallzahlen sind Rückschlüsse auf persönliche Verhältnisse der Betroffenen möglich

Die Abbildung 2 mit den erfundenen Werten zur Geburtenstatistik soll das Geschriebene noch einmal verdeutlichen. In dem Beispiel gab es 3 Drillingengeburt und eine Vierlingsgeburt. Die Vierlinge dürfen, da es

nur ein Fall ist, nicht veröffentlicht werden. Die Gefahr, dass Rückschlüsse auf die persönlichen Verhältnisse der Mutter bzw. der Eltern aus den Ergebnissen herleitbar sind, ist zu groß. Die ‚1‘ musste somit gesperrt werden. Da jedoch mit Hilfe der Summe diese ‚1‘ ausrechenbar wäre, musste eine zweite Zahl mit gesperrt werden. In einem solchen Fall würde man sich immer für die nächstgrößere Zahl entscheiden. Somit wurde im Beispiel die ‚3‘ der Drillinge mit gesperrt. Eine andere Möglichkeit wäre gewesen, und das zeigt die dritte Tabelle in Abbildung 2, statt der Sperrung der Werte die Mehrlingsgeburten zu einer Kategorie zusammenzufassen und die ‚1‘ der Vierlinge bei den Zwillingen und Drillingen mit zu „verstecken“.

4. Warum musste eine neue Methode zur Geheimhaltung bei den Bevölkerungsstatistiken eingeführt werden?

Die Frage nach dem Warum ist nicht ganz so simpel, da sowohl die bisherigen Verfahren als auch das neue Verfahren Vor- und Nachteile bei der Anwendung bieten. Einer der Gründe ist, dass es bisher mehrere Geheimhaltungsverfahren in den verschiedenen Bevölkerungsstatistiken gab (siehe Ausführungen in Abschnitt 3). Bei der einen Statistik wird gerundet, bei der anderen gesperrt. Zudem gibt es hierzu auch noch Unterschiede in den statistischen Landesämtern. Was der eine rundet, wird beim anderen gesperrt. Das ist wenig nutzerfreundlich. Eine Vereinheitlichung macht die Nutzung der Ergebnistabellen über mehrere Statistiken hinweg und im statistischen Verbund einfacher. Zudem kann man die Kommunikation vereinheitlichen.

Müssen in einer Tabelle viele kleine Fallzahlen gesperrt werden, kann der Nutzer nichts mehr mit der Tabelle anfangen

Ein weiterer Grund ist die Tatsache, dass vor allem das Verfahren der Zellsperrung keine Auswertungen mit kleinen Fallzahlen ermöglicht. Wenn ein Datennutzer zum Beispiel eine Wanderungsmatrix mit den Zu- und Fortzügen nach Altersgruppen, Geschlecht und dem Herkunfts- und Zielgebiet der Wandernden für Gemeinden bestellt, ergibt dies für Thüringen Tabellen, in denen 70 bis 80 Prozent der Felder mit nur ein oder 2 Fällen besetzt sind. Diese Felder müssen komplett gesperrt werden. Bedenkt man, dass

hinzu noch die sekundäre Sperrung weiterer Felder zur Verhinderung der Ausrechenbarkeit kommen, erkennt man schnell, dass der Datennutzer mit einer solchen Auswertung nichts anfangen kann, wenn gerade mal 10 bis 15 Prozent aller Felder besetzt sind. In einer solchen Tabelle sind somit kaum noch verwendbare Zahlen enthalten.

Ein letzter und ebenso wichtiger Grund für die Einführung einer neuen Geheimhaltungsmethode ist, dass die bisherigen Verfahren manuell angewendet wurden, das neue funktioniert automatisiert. Manuelle Arbeit ist nicht nur fehleranfällig, sie ist auch zum Teil ausgesprochen arbeitsintensiv, je nach Größe und Merkmalsvielfalt der Auswertung. Die Nachteile zeigen sich vor allem bei komplexen und sehr feingliedrigen Auswertungen wie der beschriebenen Wanderungsmatrix. Wie kommt der Datennutzer, der jedoch genau diese Auswertung benötigt, nun an verwendbare Ergebnisse der Wanderungsstatistik? An dieser Stelle hat das Cell-Key-Verfahren Vorteile.

5. Die Cell-Key-Methode – vereinfacht erklärt

Es gibt komplette Methodenaufsätze und Bücher über das Cell-Key-Verfahren. Doch für diesen Aufsatz und den verfolgten Zweck wäre es nicht zielführend, diese Geheimhaltungsmethode umfassend und detailgetreu hier zu beschreiben. Daher soll das Verfahren kurz, auf das grundlegende fokussierend, sehr vereinfacht, aber anschaulich dargestellt werden.

Nicht alle Werte einer Tabelle werden bei der Cell-Key-Methode verändert

Das Cell-Key-Verfahren ist ein automatisiertes datenveränderndes Verfahren. Mit Hilfe einer an jeden

einzelnen originalen Datensatz angespielten Zufallszahl und einer festgelegten Veränderungsmatrix wird ein Teil der originalen Werte in den Tabellenfeldern um einen vorab festgelegten Wert verändert. Die Veränderungsmatrix ist so aufgebaut, dass nur ein bestimmter Prozentsatz aller Werte tatsächlich verfälscht wird. Das führt dazu, dass im Umkehrschluss der andere Teil an Daten im Original erhalten bleibt und auch so veröffentlicht werden kann. Der Vorteil an diesem Vorgehen ist, dass niemand weiß, welcher Wert verfälscht wurde und welcher Wert noch im Original vorhanden ist. Diese Unsicherheit, die dadurch entsteht, reicht aus, um sichere Rückschlüsse auf die Angaben einzelner Personen zu verhindern. Bisher konnte eine Person, welche selbst Merkmalsträger eines mit 2 Personen besetzten Feldes ist, leicht einen Rückschluss auf die andere Person ziehen. Durch die Möglichkeit, dass der Wert in dem Feld jedoch nicht original, sondern verfälscht ist, weiß der Merkmalsträger jedoch nicht, ob es wirklich nur diese andere Person gibt. Es könnten auch 1, 3, 4, 5 oder 6 Fälle sein, je nach Verfälschungsgrad und je nach Veränderungsmatrix.

Im Folgenden soll anhand eines einfachen Beispiels das soeben Geschriebene veranschaulicht werden. Als Grundlage dient eine Beispielliste der 10 Geburten eines Beispieldorfes. Wie der Abbildung 3 zu entnehmen ist, sind zu jeder Geburt die Angaben zu Geschlecht und Nationalität enthalten. Zudem wurde jedem Datensatz eine Zufallszahl automatisch durch das Fachverfahren angespielt, was die Grundlage für das Cell-Key-Verfahren ist. Diese Zufallszahl liegt immer zwischen 0 und 1 und hat mehrere Kommastellen. In diesem Beispiel sind es zur Veranschaulichung nur 3, im Original werden dafür mehr als 10 Nachkommastellen verwendet.

Abbildung 3: Liste der Geburten im Beispieldorf

Reihenfolge der Geburt	Geschlecht des Kindes	Nationalität	Zufallszahl
1	Männlich	deutsch	0,674
2	Männlich	deutsch	0,057
3	Weiblich	deutsch	0,698
4	Männlich	nichtdeutsch	0,878
5	Weiblich	deutsch	0,598
6	Männlich	deutsch	0,289
7	Weiblich	nichtdeutsch	0,599
8	Weiblich	deutsch	0,193
9	Weiblich	deutsch	0,658
10	Weiblich	deutsch	0,726

Abbildung 4: Geburten im Beispieldorf nach Geschlecht und Nationalität (ohne Geheimhaltung)

Geschlecht	deutsch	nichtdeutsch	Summe
männlich	3	1	4
weiblich	5	1	6
Summe	8	2	10

Abbildung 5: Summe der Zufallszahlen (Cell-Keys) der Geburten im Beispieldorf nach Geschlecht und Nationalität

Geschlecht	deutsch	nichtdeutsch	Summe
männlich	1,020	0,878	1,898
weiblich	2,873	0,599	3,472
Summe	3,893	1,477	5,370

In einer einfachen Kreuzmatrix der Geburten nach Geschlecht und Nationalität des Kindes dargestellt, ergibt das die Abbildung 4.

Um den Grad der Veränderung ermitteln zu können, muss für jedes Feld nun die Summe der Zufallszahlen, der sogenannte Cell-Key, gebildet werden. Das heißt, dass die Zufallszahl von jeder Geburt, welche Bestandteil des jeweiligen Feldes ist, zusammen gerechnet wird. Das Ergebnis ist in der Abbildung 5 zu sehen.

Als zweite Grundlage benötigt man die Überlagerungsmatrix. Diese gibt vor, wie hoch der Grad der Veränderung des originalen Wertes ist. Wie hoch die Veränderung ausfällt, hängt einerseits davon ab, wie groß der originale Wert ist. Man kann somit den Grad der Verfälschung anhand der Werte in den Feldern variieren. Eine 1 wird nach dem Beispiel der abgebildeten Veränderungsmatrix nicht so stark verfälscht wie ein Wert größer 5. Zudem ist entscheidend, wo auf der Skala der Cell-Key liegt. Um dies ablesen oder ermitteln zu können, werden vom Cell-Key nur die Stellen nach dem Komma genommen.

Als Beispiel: Die Summe der Zufallszahlen der weiblichen deutschen Geburten beträgt 2,873. Die Zahl vor dem Komma wird gestrichen, es zählen nur die Nachkommastellen. Insgesamt ist das Feld zudem mit 5 Geburten besetzt. Überträgt man beide Informationen in die Überlagerungsmatrix, erkennt man, dass die Zahl um einen Wert von -2 verändert werden muss. Ein weiteres Beispiel: Beim Insgesamt-Ergebnis der weiblichen Geburten beträgt die Summe der Zufallszahlen 3,472. Die Zahl vor dem Komma wird wieder ignoriert. Es sind zudem insgesamt 6 Fälle. Aus der Überlagerungsmatrix kann man nun ablesen, dass dieser Wert nicht verfälscht wird.

In Abbildung 7 ist die Geheimhaltung an der Beispieltabelle mit der Cell-Key-Methode vollständig durchgeführt worden. Es wurden nur 4 von 9 Werten verändert. Jedoch ist sofort erkennbar, was einer der großen Nachteile der Cell-Key-Methode ist: Die Werte sind nicht mehr aufrechenbar, da die Summen eine eigene Zufallszahl haben und ebenso verändert werden. Dennoch bleiben 5 von 9 Werten im Original erhalten. Jedoch kann kein außenstehender Datennutzer wissen, welche Werte dies sind. Veröffentlicht oder für Datennutzer zur Verfügung gestellt würde

Abbildung 6: Beispieltabelle Überlagerungsmatrix

Feldwert	0,00 - 0,09	0,10 - 0,19	0,20 - 0,29	0,30 - 0,39	0,40 - 0,49	0,50 - 0,59	0,60 - 0,69	0,70 - 0,79	0,80 - 0,89	0,90 - 0,99
5+										
4										
3										
2										
1										
0										

Cell Key

Veränderungsgrad Legende: +3 +2 +1 +/- 0 -1 -2 -3

Abbildung 7: Ergebnis der Überlagerung bei der Beispieltabelle

Geburten im Beispieldorf nach Geschlecht und Nationalität (mit Geheimhaltung - ckm)

Geschlecht	deutsch	nichtdeutsch	Summe
männlich	5	0	3
weiblich	3	1	6
Summe	6	2	10

nur diese letzte überlagerte Tabelle. Die Tabelle mit den originalen Werten sowie die Überlagerungsmatrix werden niemandem zur Verfügung gestellt.

Auch Tabellen mit kleinen Fallzahlen können mit Hilfe der Cell-Key-Methode ausgewertet und veröffentlicht werden

Der große Vorteil der Cell-Key-Methode ist, dass Werte im Original erhalten bleiben und dass die Tabelle auch mit kleinen Fallzahlen auswertbar ist. Nach dem vorherigen Verfahren, der Zellspernung, hätte diese Beispieltabelle nicht veröffentlicht werden können. Die beiden „1“ und die „2“ in der Summe bei den nichtdeutschen Geburten hätten gesperrt werden müssen. Durch die sekundäre Sperrung hätten auch die deutschen Geburten gesperrt werden müssen. In der Folge wäre die Abbildung 4 nahezu leer gewesen, nur die Werte in der Spalte „Summe“ hätten veröffentlicht werden können. Der potentielle Datennutzer hätte mit der Tabelle nichts anfangen können. Mit der Cell-Key-Methode bleiben jedoch alle Felder besetzt.

Ein zweiter großer Nachteil neben der Nichtaddierbarkeit dieser Methode ist, dass sich errechnete Werte, wie beispielsweise ein natürlicher Saldo (Geburten minus Sterbefälle) oder ein Wanderungssaldo (Zuzüge minus Fortzüge) ins Gegenteil umkehren können. Folgendes Beispiel: Eine Gemeinde konnte im Original in einem Berichtsjahr 4 Zuzüge und 5 Fortzüge registrieren. Das entspricht einem Saldo von -1. Die Gemeinde hat also einen Einwohner durch den Fortzugsüberschuss verloren. In den Ergebnistabellen der Wanderungsstatistik kann durch die Verfälschung durch Anwendung der Cell-Key-Methode aus dem Wanderungsverlust ein Wanderungsgewinn werden. Wenn beispielsweise die Zufallszahlen dazu führen, dass die 4 Zuzüge zu 6 Zuzügen verändert werden, und die Fortzüge unverändert bleiben oder gar um einen Fortzug weniger auf 4 verfälscht werden, hat die Gemeinde statt des

echten Verlustes nun ein Wanderungsplus in Höhe von 1 bis 3 Personen in den Ergebnistabellen der Wanderungsstatistik stehen. Dies ist alles abhängig von der Zufallszahl und dem Grad der Veränderung der originalen Werte. Fakt ist jedoch auch, dass je kleiner die Besetzungszahlen in den Tabellenfeldern sind, umso größer wirkt sich die Verfälschung aus.

6. Auswirkungen auf die Ergebnisdarstellung der Bevölkerungsstatistiken

Die Statistischen Landesämter haben gemeinsam mit dem Statistischen Bundesamt beschlossen, dass die Ergebnisse der betroffenen Bevölkerungsstatistiken³⁾ ab dem Berichtsjahr 2025 ausschließlich nur noch mit den durch die Cell-Key-Methode überlagerten Ergebnissen veröffentlicht werden dürfen. Das betrifft alle Ergebnisse, auch die vorläufigen Monats- und Quartalszahlen. Alle Tabellen und Auswertungen in sämtlichen Produkten und Darstellungsformen, seien es Faltblätter, Statistische Berichte, Tabellen in Auskunftsdatenbanken, Pressemitteilung, Nutzeranfragen oder Forschungsanfragen, alle werden ab 2025 nur noch mit überlagerten Ergebnissen veröffentlicht. Rückwirkend werden die Ergebnisse in Thüringen jedoch nicht verändert. Alles, was bisher in den Bevölkerungsstatistiken veröffentlicht wurde, wird nicht nachträglich überlagert. Das heißt, dass es sowohl in sämtlichen Veröffentlichungsprodukten als auch in den Datenbereitstellungen für Datennutzer der Bevölkerungsstatistiken einen Bruch in den Zeitreihen geben wird. Bis 2024 werden alle Ergebnisse noch nach dem bisherigen Vorgehen veröffentlicht, ab 2025 mit den durch die Cell-Key-Methode teils veränderten Ergebnissen.

Datennutzer müssen ihre Anfragen so konkret wie möglich im Vorfeld präzisieren

Die oben beschriebenen Nachteile der Cell-Key-Methode, zum Beispiel die nicht vorhandene Additivität

3) Statistik der Einbürgerungen, Statistik der Geburten, Statistik der Sterbefälle, Statistik der Eheschließungen, Wanderungsstatistik, Todesursachenstatistik

der Werte einer Tabelle, wird der Nutzer der Bevölkerungsstatistiken sehen und spüren. Dies wird sich vor allem dann als problematisch erweisen, wenn der Datennutzer im Nachgang der Datenbereitstellung selbstständig weitere Aggregation der Zahlen oder Werte durchführen möchte. Dies ist im Grunde nicht oder nur mit Einschränkung möglich, da die Ergebnisse noch stärker verfälscht werden, als sie es schon sind. Der Datennutzer muss somit bereits bei der Zusammenstellung seiner Anfrage die Merkmalsausprägungen so präzise wie möglich definieren.

Nach den getroffenen Festlegungen werden jedoch errechnete Indikatoren und Werte auch von den Statistischen Ämtern mit überlagerten Werten berechnet. Das betrifft beispielweise Maßzahlen wie die Geburtenziffer oder die Sterbefälle je 1000 Einwohner. Somit ist gewährleistet, dass alle Datennutzer auf dieselben errechneten Werte kommen, da den außenstehenden Datennutzern die originalen unveränderten Werte nicht zur Verfügung stehen. Zudem werden Salden mit veränderten Werten berechnet, nicht mit den originalen Werten.

Die veröffentlichte Bevölkerungsbilanz ist nicht mehr nachrechenbar

Obwohl anfangs erwähnt wurde, dass die Statistik der Bevölkerungsfortschreibung nicht von dieser Methode der Geheimhaltung betroffen ist, stimmt diese Aussage zwar grundsätzlich, aber auch wieder nicht. Denn eine häufig genutzte und angefragte Tabelle ist

die Bevölkerungsbilanz der Gemeinden, Kreise und des Freistaats Thüringen. Die Bilanzen setzen sich in der Regel aus dem Bevölkerungsanfangsbestand, den hinzuzurechnenden Geburten und Zuzügen sowie den abzuziehenden Sterbefällen und Fortzügen zusammen. Diese Bevölkerungsbewegungen ergeben die Bevölkerungszahl am Jahresende. Im Hintergrund werden die Bevölkerungszahlen auch mit den originalen, unverfälschten Werten ermittelt. Aufgrund der Tatsache, dass die Veröffentlichung der genannten Bewegungen ab 2025 der Verfälschung durch die Cell-Key-Methode unterliegen, müssen die Werte in den Bilanzen aber auch verfälscht werden. Außenstehende können somit zukünftig die Bilanz der Bevölkerungsbewegung und der Entstehung der Bevölkerungszahlen nicht mehr nachrechnen.

Umstiegsprozess auf die Cell-Key-Methode muss kommunikativ begleitet werden

Die Einführung der Cell-Key-Methode wird nicht bei allen Datennutzern auf Verständnis stoßen. Es ist daher essentiell, dass der Umstiegsprozess intensiv durch den zuständigen Fachbereich des Thüringer Landesamtes für Statistik kommunikativ begleitet wird. Dies wird durch persönliche Beratung, aber auch mit Hilfe von Sprachregelungen und Methodenpapieren für interessierte Nutzerinnen und Nutzer erfolgen.