

Neues Geheimhaltungsverfahren des Statistischen Landesamtes

Teil 1: Aspekte der Statistischen Geheimhaltung

Von Harald Wirtz und Claudia Baier

Die Gewährleistung der Geheimhaltung ist eine fundamentale Aufgabe der statistischen Ämter des Bundes und der Länder. Durch die Geheimhaltung wird die für die Aussagefähigkeit der Daten unabdingbare Vertrauensbasis geschaffen und erhalten. Seit 1987 wird vor diesem Hintergrund im Statistischen Landesamt bei der Durchführung der Geheimhaltungsprüfungen ein großrechnerbasiertes Verfahren eingesetzt. Wegen seiner mangelnden Integration in den Aufbereitungsprozess des Landesinformationssystems (LIS) und der im Verbund zu Ende gehenden Großrechner-Ära wurden im Jahr 2009 die Konzeption und Programmierung eines neuen Geheimhaltungsverfahrens in Angriff genommen. Das neue Verfahren steht zwischenzeitlich für den Produktiveinsatz bereit.

Einführung

Geheimhaltungsprüfungen werden im StLA Rheinland-Pfalz maschinell durchgeführt

Das Statistische Landesamt Rheinland-Pfalz unterhält seit Anfang der 70er-Jahre ein datenbankgestütztes Landesinformationssystem (LIS). Ziel dieses Systems ist die flexible und an adhoc-Bedürfnissen ausgerichtete Bereitstellung von Daten der amtlichen Statistik. Die erste Version des LIS basierte auf der Großrechnertechnologie und auf einer hierarchischen Datenbank. Seit 2003 wird ein neues LIS eingesetzt. Dieses wurde – der informationstechnologischen Entwicklung folgend – als Client-Server-Anwendung konzipiert und nutzt die Funktionalitäten einer relationalen Datenbank. Im LIS stehen rund 130 Millionen Daten für Auswertungen zur Verfügung. Parallel zum Betrieb und zur Weiterentwicklung des LIS kam im

Statistischen Landesamt der Realisierung von maschinellen Verfahren zur Gewährleistung der statischen Geheimhaltung eine hohe Priorität zu. Seit 1987 wird ein großrechnerbasiertes Zellsperrverfahren bei der Durchführung der primären und der sekundären Geheimhaltungsprüfungen verwendet. Die hierbei ermittelten Sperrmuster wurden zum einen bei der Einlagerung in die LIS-Datenbank, zum anderen im Rahmen von sonstigen Tabellierungsprogrammen genutzt. Aufgrund der im statistischen Verbund zu Ende gehenden Großrechner-Ära und vor dem Hintergrund der Zielsetzung „Schaffung eines effizienten Workflows zur Abwicklung der Geheimhaltungsprüfungen“ wurden im Jahr 2009 die Konzeption und Programmierung eines neuen Geheimhal-

Ü 1

Eckpunkte der Geheimhaltung

Was muss geheim gehalten werden?	Normen/Regeln ↓	<ul style="list-style-type: none"> ■ § 16 BStatG ■ Spezialgesetzliche Regeln ■ ...
Wie wird geheim gehalten?	Methoden ↓	<ul style="list-style-type: none"> ■ Datenreduzierung ■ Datenveränderung ■ ...
Womit wird die Geheimhaltung umgesetzt?	Verfahren ↓	<ul style="list-style-type: none"> ■ Computerprogramme ■ Manuell ■ ...
Wann erfolgt die Geheimhaltungsprüfung?	Zeitpunkt	<ul style="list-style-type: none"> ■ Dateieinlagerung ■ Datenabruf

tungsverfahrens in Angriff genommen. Die zwischenzeitlich weitestgehend abgeschlossene programmiertechnische Umsetzung zeichnet sich durch die Schlagworte „Datenbankbasierung“, „automatisierter Workflow“ und „integrative Anbindung an das Auskunftssystem des LIS“ aus.

Aspekte der statistischen Geheimhaltung

Geheimhaltung schafft Vertrauen

Die Gewährleistung der Geheimhaltung ist eine fundamentale Aufgabe der statistischen Ämter des Bundes und der Länder. Durch die Geheimhaltung wird die für die Aussagefähigkeit der Daten unabdingbare Vertrauensbasis geschaffen und erhalten. Andererseits ist mit dem Schutz geheim zu haltender Daten gegen deren Offenlegung ein Informationsverlust verbunden, der die Aussagefähigkeit der veröffentlichten Statistiken einschränkt.¹ Die im Rahmen der Durchführung der Geheimhaltungsprüfung zu lösende Optimierungsaufgabe besteht somit darin, mit geeigneten und einheitlich zur Anwendung kommenden Methoden das

Offenlegungsrisiko von Daten zu minimieren, gleichzeitig aber den Informationsgehalt veröffentlichter Tabellen so hoch wie möglich zu halten. Dies vor dem Hintergrund, dass die strikte Geheimhaltung der zu statistischen Zwecken erhobenen Einzelangaben nicht nur konstitutiv für die Funktionsfähigkeit der amtlichen Statistik, sondern auch im Hinblick auf das Recht auf informationelle Selbstbestimmung unverzichtbar ist.

Die Sicherstellung der statistischen Geheimhaltung stellt eine äußerst komplexe, facettenreiche und arbeitsintensive Aufgabe dar. Bevor das neue Geheimhaltungsverfahren des LIS näher beschrieben wird, müssen deshalb zunächst einige rechtliche, methodische und verfahrenstechnische Aspekte der statistischen Geheimhaltung betrachtet werden. Hierdurch werden die Rahmenbedingungen aufgezeigt, die bei der Entwicklung des neuen Verfahrens zu beachten waren.

Der Schwerpunkt der Ausführungen liegt nachfolgend auf dem zweit- und dem drittgenannten Aspekt. Es wird gezeigt, in welchen ablauforganisatorischen Teilschritten die Geheimhaltungsprüfungen durchgeführt

¹ Vgl. Sarah Giessing und Stefan Dittrich, Tabellengeheimhaltung im statistischen Verbund – ein Verfahrenvergleich am Beispiel der Umsatzsteuerstatistik, Statistisches Bundesamt, Wirtschaft und Statistik 8/2006, S. 805

² BT-Drucksache Nr. 10/5345 vom 17. April 1986.

werden, welche Methoden (wahlweise) zur Anwendung kommen und welche Dokumentations-/Protokollierungsmechanismen zur Verfügung stehen.

Rechtliche Aspekte

§ 16 BStatG regelt die statistische Geheimhaltung

Die statistische Geheimhaltung ist in § 16 des Gesetzes über die Statistik für Bundeszwecke (BStatG) geregelt. Danach sind „Einzelangaben über persönliche und sachliche Verhältnisse, die für eine Bundesstatistik gemacht werden [...] geheim zu halten, soweit durch Rechtsvorschrift nichts anderes bestimmt ist“. In der Begründung zum BStatG heißt es: „Die Geheimhaltung ist seit jeher das Fundament der Bundesstatistik. Ihre Gewährleistung dient [...] folgenden Zielen:

- Schutz des Einzelnen vor der Offenlegung seiner persönlichen und sachlichen Verhältnisse,
- Erhaltung des Vertrauensverhältnisses zwischen den Befragten und den statistischen Ämtern,
- Gewährleistung der Zuverlässigkeit der Angaben und der Berichtswilligkeit der Befragten.“

Als Ausnahmen von dem generellen Geheimhaltungsgebot nennt § 16 BStatG unter anderem folgende Fälle:

- Die Veröffentlichung oder Übermittlung erfolgt als statistisches Ergebnis in zusammengefasster Form („Tabelle“).
- Die Einzelangabe ist dem Befragten oder Betroffenen nicht zuzuordnen.
- Der Befragte hat in die Übermittlung oder Veröffentlichung schriftlich eingewilligt.

Für die praktische Ausgestaltung der Geheimhaltungsprüfungen sind in diesem

³ Zusatzwissen (= Vorwissen) beinhaltet solche Erkenntnisse, die ein Nutzer unabhängig von der Veröffentlichung einer Tabelle hat und anhand derer die Eingrenzung von Tabellenwerten vorgenommen werden kann.

Zusammenhang folgende Punkte von allgemeiner Bedeutung:

- Die Sensibilität einer Einzelangabe spielt keine Rolle: Die Diagnose „Darmkrebs“ und die Diagnose „Herzinfarkt“ sind beide gleichermaßen geheim zu halten.
- Die näherungsweise Offenlegung einer Einzelangabe darf nicht möglich sein.
- Die Angaben über eine Person, ein Unternehmen usw. sind auch dann geheim zu halten, wenn die gleichen Informationen öffentlich zugänglich sind.

Aus § 16 BStatG lässt sich nicht unmittelbar ableiten, was bezüglich der Geheimhaltung im Einzelfall zu tun ist. Ob eine Tabelle als absolut anonym gelten kann oder nicht, hängt von mehreren Faktoren ab, z. B. den verfügbaren Informationen oder dem bei einem potenziellen Datenangreifer unterstellten (Zusatz-) Wissen³. Das bei einzelnen Personen oder durch den Zusammenschluss von Personen möglicherweise entstehende Zusatzwissen ist letztendlich unbekannt. Vor diesem Hintergrund besteht das Ziel der Geheimhaltung darin, dass zufällige Einsichten unwahrscheinlich werden und dass gezielte Einsichten nur mit einem äußerst hohen Aufwand möglich sind.

Methodische Aspekte

In methodischer Hinsicht beschränken sich die nachfolgenden Ausführungen zur LIS-Geheimhaltung auf die Ergebnisdarstellung in statistischen Tabellen. Tabellen sind aufgrund ihrer Kompaktheit, ihres strukturierten Aufbaus und der Möglichkeit, durch tiefe Schachtelung komplexe Berichte zu erstellen, nach wie vor ein zentrales Instrument zur Präsentation statistischer Informationen. Gemäß DIN 55 301 sind (Einzel-)Tabellen strukturierte Zusammenstellungen von Tex-

Statistische Tabellen nach wie vor ein zentrales Mittel zur Präsentation von Statistiken

ten und/oder statistischen Daten (Tabelle = „Kreuzkombination von Gliederungsmerkmalen“). Die darzustellenden Inhalte werden in Zeilen und Spalten eingetragen, die aneinander ausgerichtet werden. Man spricht von n-dimensionalen Tabellen, wenn diese nach n verschiedenen Merkmalen klassifiziert sind und jeweils nur eine Randsumme für jedes Klassifizierungskriterium aufweisen. Übergreifende Tabellen verfügen über eine Menge gemeinsamer Tabellenwerte, weisen also Schnittmengen auf.

Eine weitere Differenzierung ergibt sich aus der Unterscheidung zwischen Mikro- und Makrodaten bzw. Mikro- und Makrotabellen. Mikrodaten entsprechen den Einzelbeobachtungen; sie sind das Ergebnis der statistischen Erfassungs- und Aufbereitungsphase. Makrodaten hingegen stehen für eine auswertungsorientierte Zusammenfassung von Mikrodaten und dienen somit dem Ziel der Auswertungsphase. Ergänzend zu den Mikrodaten und Makrodaten sind die Metadaten zu sehen. Im Sinne von Konstruktionsdaten beschreiben sie vollständig und widerspruchsfrei die Menge der betrachteten Mikro- und Makrodaten. Der Übergang von Mikrodaten zu Makrodaten durchläuft im Rahmen des statistischen Aufbereitungsprozesses mehrere Verarbeitungsstufen. Vor dem Hintergrund der statistischen Geheimhaltung ist von Bedeutung, dass bei fachlich und/oder regional stark gegliederten Tabellen die Situation auftreten kann, dass einzelfallbezogene Daten (= Mikrodaten) sich unmittelbar in einer tabellarischen Auswertung (= Makrodaten) wieder finden können.

⁴ In methodischer Hinsicht wird eine Reduzierung auch dann erreicht, wenn die Klassenbreiten einer Auswertung vergrößert werden, sodass mehr Einzelwerte in den Klassen und damit in den in der Tabelle dargestellten Makrodaten enthalten sind. Die gleiche Wirkung kann durch eine Zusammenfassung von Klassen erzielt werden.

⁵ Vgl. Nils Radmacher-Nottelmann, Geheimhaltung mit Makrodaten – Das Beispiel der Beherbergungsstatistik, Statistische Analysen und Studien NRW, Band 19, Seite 32.

Geheimhaltung durch Datenreduzierung oder Datenveränderung

In Makrotabellen werden die Einzeldaten von Merkmalsträgern – wobei die zurückliegend genannte Einschränkung zu beachten ist – zu gemeinsamen Positionen zusammengefasst. Mikrotabellen hingegen geben jeweils die Individualdaten der Auskunftgebenden wieder. Einzelne in einer Makrotabelle veröffentlichte Werte können geheim gehalten werden, in dem die kritischen Angaben entweder reduziert oder verändert werden. Ersteres kann z. B. dadurch erreicht werden, dass die geheim zu haltenden Tabellenzellen vollständig gesperrt, d. h. von der Anzeige ausgeblendet werden.⁴ Ein Nachteil dieser Zellsperre ist allerdings, dass der gesamte betroffene Wert verdeckt wird und damit viel Information verloren geht. Alternativ können die Tabellenwerte – z. B. nach den Regeln eines Rundungsalgorithmus – verändert und anschließend veröffentlicht werden.⁵ Die veränderten Werte geben dann allerdings die von der Statistik abgebildete Realität nicht mehr korrekt wieder.

Zellsperre verdeckt; Datenveränderung verschleiert

Geheimhaltungsprüfungen im Workflow von Informationssystemen

Das in den Workflow des Landesinformationssystems integrierte Geheimhaltungsverfahren arbeitet nach der Methode der Zellsperre. Vor dem Hintergrund dieser Festlegung wurden im Vorfeld der programmiertechnischen Umsetzung zwei Alternativen der ablauforganisatorischen Ausgestaltung geprüft:

- Alternative 1: Die im Rahmen von Datenbankabfragen erstellten Tabellierungen greifen unmittelbar auf die Mikrodaten der jeweiligen Statistik zu, werden realtime,

d. h. in Echtzeit, unter Geheimhaltungsaspekten qualifiziert und im Rahmen der Bildschirmpräsentation bzw. des Ausdrucks mit entsprechenden Sperrmustern versehen.

- Alternative 2: Das in die Auskunftsdatenbank aufzunehmende Datenmaterial wird zunächst in der durch seine Klassifizierung vorgegebenen Granularität in einem separaten Arbeitsschritt in geheimhaltungsmäßiger Hinsicht qualifiziert. Hierbei werden die geheim zu haltenden Merkmale auf der Makroebene um spezielle Geheimhaltungskennzeichen ergänzt. Sie werden anschließend zusammen mit den zugehörigen Merkmalen in die Datenbank eingelagert. Die so vorgenommene Merkmalsqualifizierung wird zum Zeitpunkt des Datenabrufs bzw. der Tabellierung dazu genutzt, die notwendigen Zellsperren in den jeweiligen Ergebnisdarstellungen unterschiedlichster Aggregationsstufen vorzunehmen.

Die zuerst genannte Vorgehensweise war unter Performance-Gesichtspunkten nicht realisierbar. Darüber hinaus müsste ein derartiges Verfahren die im Rahmen von flexiblen Auswertungen jeweils ermittelten Sperrmuster speichern und bei zukünftigen (anders strukturierten) Auswertungen im Sinne einer tabellenübergreifenden Geheimhaltung berücksichtigen; eine in der Realität nicht zu erfüllende Anforderung. Im Rahmen der Entwicklung des in den Workflow des LIS zu integrierenden Geheimhaltungsverfahrens wurde deshalb das zweitgenannte Vorgehen favorisiert.

⁶ Die Einteilung (Klassifizierung) von Objekten in eine Hierarchie impliziert eine Rangordnung, nach der Datenobjekte geordnet werden können. Die einzelne Stufe einer Hierarchie/Klassifikation beschreibt den jeweiligen Verdichtungsgrad der Daten. Bei einer einfachen Hierarchie enthält die höhere Hierarchiestufe die aggregierten Werte genau einer niedrigeren Stufe. Eine Hierarchie kann 1 bis n Stufen beinhalten. Hierarchische Klassifikationen in der amtlichen Statistik sind z. B. der AGS oder die Wirtschaftszweigsystematik.

Die im LIS in n-dimensionalen Quadern zu speichernden bzw. sonstigen Tabellierungsprogrammen zuzuführenden Daten werden in einem der Einlagerung vorgeschalteten Arbeitsschritt einer systematischen Geheimhaltungsprüfung unterzogen. Dabei werden in einer ersten Phase die geheimhaltungskritischen Tabellenfelder, die von der Anzeige auszuschließen sind, identifiziert und auf der Makroebene entsprechend „gekennzeichnet“ (primäre Geheimhaltung). Diese Felder werden in späteren LIS-Abrufen bzw. Veröffentlichungen zwingend von der Anzeige ausgeschlossen. Um zu gewährleisten, dass diese primär geheimen Felder nicht durch Differenzbildung zu ebenfalls veröffentlichten Rand- und Zwischensummen aufgedeckt werden können, sind in einem zweiten Arbeitsgang zusätzliche Merkmals- und/oder Summenfelder „gegen-“zusperrn. In die entsprechende Prüfung (sekundäre Geheimhaltung) werden enumerativ alle Rand- und Zwischensummen des n-dimensionalen Datenraums einbezogen. Dies ist Voraussetzung dafür, dass die in das LIS integrierten OLAP-Komponenten auch bei geheimhaltungskritischen Daten genutzt werden können. Die geheimhaltungstechnische Qualifizierung (Kennzeichnung) erfolgt zugleich so, dass Daten in OLAP-Auswertungen (z. B. den Pivot-Tabellierungen des LIS) dynamisch wieder ab der Aggregationsstufe angezeigt werden, ab der keine Geheimhaltungskritikalität mehr besteht. Unabhängig hiervon wird im Rahmen des LIS die Notwendigkeit der tabellenübergreifenden Geheimhaltung durch ein entsprechendes Quaderdesign minimiert: Die im LIS gespeicherten geheimhaltungsqualifizierten Quader weisen als n-dimensionale Tabellen bis zu sieben Hierarchien⁶ und damit eine entsprechend niedrige Granularität auf.

Primäre und sekundäre Geheimhaltung: zwei Säulen der Zellsperrenverfahren

Methoden der primären und sekundären Geheimhaltung

Primäre Geheimhaltung dient der Vermeidung exakter und näherungsweise Offenlegungen

Primäre Geheimhaltungsmethoden, die im Rahmen der Zellsperverfahren eine exakte Offenlegung geheimhaltungskritischer Tabellenwerte verhindern, werden als Fallzahlenregeln bezeichnet.⁷ Hiernach ist ein Wert geheim zu halten, wenn (auf Mikrodatenebene) eine bestimmte Mindestfallzahl, der zu diesem Tabellenfeld beitragenden Einzelangaben unterschritten wird.⁸ Die Anzahl der zum Wert einer Zelle beitragenden Einzelangaben wird mit n bezeichnet. Die Fallzahlenregel könnte z. B. lauten: „Ein Tabellenwert ist geheim zu halten, wenn weniger als drei Befragte (Einheiten) zu dem Wert beigetragen haben.“ Fallzahlenregeln sind allerdings nicht mehr ausreichend, wenn eine näherungsweise Offenlegung von Einzelangaben ausgeschlossen werden soll. Eine näherungsweise Offenlegung ist dann möglich, wenn ein Tabellenwert von einer Einzelangabe (Mikrodatum) dominiert wird. Zur Absicherung solcher Zellenwerte wurden die sogenannten Dominanzregeln entwickelt.

- n,k -Dominanzregeln: Ein Tabellenwert X ist dann geheim zu halten, wenn (auf der Mikrodatenebene) der Anteil der n größten Einzelwerte mehr als k Prozent beträgt. Für $n = 2$ ergibt sich zum Beispiel folgende Formel:

$$x_1 + x_2 > \frac{k}{100} \cdot X$$

Bei dieser $2,k$ -Dominanzregel wird also berücksichtigt, dass es Datenangreifer

gibt, die zumindest einen Beitrag zum Gesamtwert, nämlich ihren eigenen, kennen.⁹ Wird ein Tabellenwert z. B. durch die beiden größten Einzelangaben dominiert, können diese mithilfe des Aggregatwertes des Tabellenfeldes relativ genau nach oben abgeschätzt werden. Am präzisesten kann hierbei der größte Einzelwert eines Tabellenfeldes geschätzt werden, und zwar durch den Befragten mit dem zweitgrößten Einzelwert, indem dieser seinen Wert vom Aggregatwert des Tabellenfeldes subtrahiert.

Der Nachteil der $2,k$ - bzw. der n,k -Dominanzregeln besteht aber darin, dass eine hohe Zahl von Tabellenwerten als vertraulich eingestuft wird.

- $p\%$ -Dominanzregel: Nach dieser Regel ist der Wert eines Tabellenfeldes dann geheim zu halten, wenn er nach Abzug des zweitgrößten Einzelwerts den größten Einzelwert um weniger als $p\%$ übersteigt:

$$\frac{(X - x_2) - x_1}{x_1} \cdot 100 < p$$

Der Befragte mit dem zweitgrößten Beitrag zum Tabellenwert soll den Wert des Meistbeitragenden also um mindestens $p\%$ überschätzen, wenn er die Differenz „Gesamtwert minus eigener Wert“ heranzieht. Der Schwellenwert für p ist vorab festzulegen (z. B. $p = 8$). Die $p\%$ -Regel bezieht somit ebenfalls das mögliche Wissen des Zweitgrößten in die Betrachtung ein. Sie bietet den Vorteil, dass bei gleichem Sicherheitsniveau in der Regel weniger vertrauliche Zellen als bei der $2,k$ -Dominanzregel ermittelt werden. Der Schwellenwert für p ist für jede Statistik vorab individuell festzulegen.¹⁰ Je stärker der größte Einzelwert (x_1) geschützt werden soll, um so höhere Werte müssen für p angesetzt werden. Der letztendlich

7 Vgl. Rüdiger Dietz Repsilber, Sicherung persönlicher Angaben in Tabellendaten, Statistische Analysen und Studien NRW 1/2002, S. 25.

8 Vgl. Pierre Wettig, Beschreibung von Verfahren zur statistischen Geheimhaltung in Tabellen und ihre Anwendung, Diplom-Arbeit an der Technischen Universität Illmenau, Fakultät für Wirtschaftswissenschaften und Informatik, Februar 2002, S. 6 ff.

9 Hierbei ist es unerheblich, ob der Nutzer, der die Schätzung durchführt, sich der Genauigkeit seines Ergebnisses bewusst ist.

10 Die kritische Schwelle ist relativ schwer zu ermitteln, da der Informationsgewinn, den der Leser eines Tabellenwertes erfahren kann, in hohem Maße von den Informationen abhängig ist, die ihm neben der Statistik zur Verfügung stehen, z. B. aufgrund lokaler Ortskenntnisse.

bei einer Statistik eingesetzte Parameter unterliegt ebenfalls der Geheimhaltung, damit möglichst wenig Informationen über die Zusammensetzung des „Rests“ bekannt sind.¹¹

Im Jahr 2001 beschlossen die Statistischen Ämter des Bundes und der Länder einer Empfehlung des Arbeitskreises für Fragen der mathematischen Methodik folgend,

- die Fallzahlenregel in Kombination mit der 1,k-Dominanzregel nicht mehr anzuwenden, da sie nicht geeignet sind, die Offenlegung der zum Wert einer Tabellenzelle beitragenden Einzelangaben auszuschließen und
- diese durch die p%-Regel oder die 2,k-Dominanzregel zu ersetzen, wobei der p%-Regel der Vorzug zu geben ist.

Das in den Workflow des Landesinformationssystems integrierte Geheimhaltungsverfahren erlaubt hinsichtlich der primären Geheimhaltung die 2,k- und die p%-Dominanzregel – gesteuert über eine entsprechende Menü-Funktion – zu verwenden. Hierbei können beide Geheimhaltungsregeln parallel zur Anwendung gebracht werden. Die Protokollierung der jeweils vergebenen Geheimhaltungskennzeichen erlaubt detaillierte Analysen zur Beantwortung der Frage, welcher Regel der Vorzug zu geben ist.

Die Regeln für die primäre Geheimhaltung beschränken das Risiko der näherungsweise Offenlegung von Einzelangaben. Sie dienen dazu, solche Zellenwerte zu finden, bei denen die Einzelbeiträge auf nicht mehr akzeptable Weise geschätzt werden können. In diesem Zusammenhang ist zu bedenken, dass aggregierte Daten in Tabellen mit Zwischen- und/oder Randsummen in einem additiven Zusammenhang stehen. Um eine

¹¹ Siehe Fußnote 5.

Definitionen

Im Rahmen von **Client-Server-Anwendungen** werden informationstechnische Aufgaben und Dienste in einem Netzwerk verteilt. Einzelne Aufgaben werden von Programmen/Computern erledigt, die in Clients und Server unterschieden werden können. Clients (Dienstenutzer), z. B. der PC am Arbeitsplatz, nehmen Aufgaben des Servers (Diensteanbieters) in Anspruch, z. B. den Abruf von Daten aus einer zentralen Datenbank.

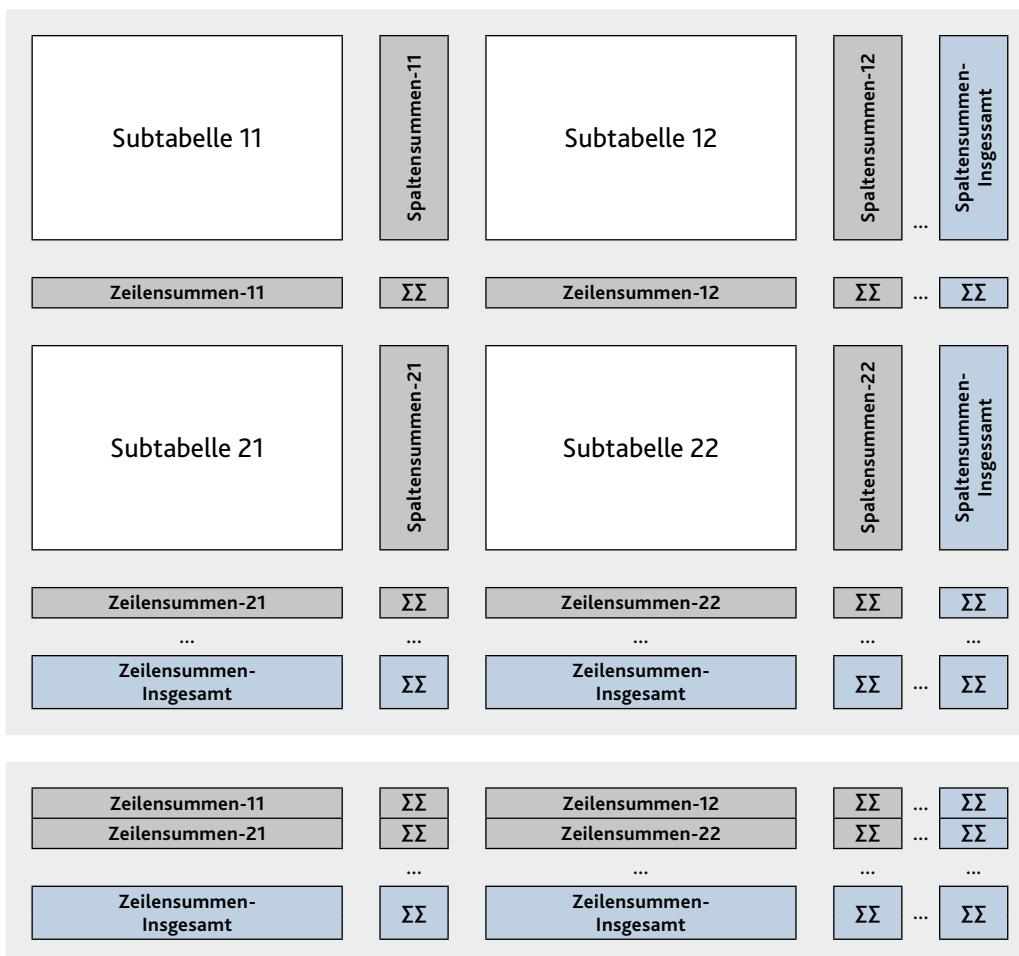
Die **Granularität** ist ein Maß für die Feinkörnigkeit (= regionale und/oder fachliche Tiefengliederung) der Daten. Sie ist damit zugleich ein Maß für die „Passgenauigkeit“ einer Information aus Sicht des Empfängers. Daten, die wenig detailliert sind, weisen eine hohe Granularität auf und umgekehrt.

Die **Kritikalität** einer Einheit drückt aus, welche Bedeutung ihrem Fehlverhalten beigemessen wird. Die Kritikalität wird in Stufen angegeben, wobei die Einstufung umso höher ist, je gravierendere Auswirkungen bei Fehlverhalten zu erwarten sind.

On Line Analytical Processing (OLAP): Technik der multidimensionalen Kreuztabellierung von Daten zum Zwecke der Auswertung, Analyse und Präsentation. Typische OLAP-Funktionalitäten sind das Tauschen der Spalten- und Zeilendimensionen (Pivotieren), das Wechseln von einer höheren zu einer niedrigeren Aggregationsstufe und umgekehrt (Drill Down/Drill Up) oder das flexible „Ausschneiden/Einfügen“ von Dimensionen/Klassen (Slicing).

Ü 2

Subtabelle – Aggregattabellen



Rückrechenbarkeit der primär gesperrten Daten durch Summen- oder Differenzbildung zu verhindern, müssen deshalb zusätzlich sogenannte Sekundärsperren vorgenommen werden. Diese Notwendigkeit besteht im Hinblick auf das Tabelleninnere einzelner Subtabellen, deren Randsummen und das aus den Randsummen ableitbare System von Aggregattabellen.

Mehrdimensionale statistische Tabellen (analog: mehrdimensionale Datenquader) können als ein System strukturierter Subtabellen verstanden werden, das – im Idealfall – in horizontaler und in vertikaler Hinsicht durchgängig summierbar ist.

Wird eine mehrdimensionale Tabelle durch das Ausblenden des Tabelleninneren der Subtabellen ST_{ij} verdichtet, so ergibt sich ein System entsprechender Aggregattabellen.

Im Rahmen einer umfassenden Sicherstellung der Geheimhaltung muss die Aufdeckung kritischer Daten für alle Subtabellen und für jede Form der Aggregattabellen eines n-dimensionalen Datenraums ausgeschlossen sein. Diese Anforderung gilt insbesondere für das OLAP-basierte Auswertungssystem des LIS, das eine dynamische Pivotierung des Datenmaterials und eine flexible Generierung tabellarischer Auswertungen erlaubt.

Sekundäre
Geheimhaltung
durch iteratives
Setzen von
Zellsperren

Der in den Workflow des Landesinformationssystem integrierte sekundäre Geheimhaltungsalgorithmus prüft alle Summanden des durch die Quaderstruktur vorgegebenen und der primären Geheimhaltungsprüfung bereits unterzogenen n-dimensionalen Datenraums enumerativ gegen alle Summierungen dieses Datenraums. Hierbei werden im Bedarfsfall sogenannte Gegensperren gesetzt. Diese können die Summanden als auch die Summenwerte betreffen. In der Regel gelingt es nicht, die sekundäre Geheimhaltungsqualifizierung in einem Durchlauf abzuschließen. Vielmehr ist ein iteratives Vorgehen erforderlich; neu gesetzte Gegensperren erfordern ggf. weitere Gegensperren. Die aktuell an den Beispielen „Umsatzsteuerstatistik (Vorankmeldungen)“ und „Landwirtschaftszählung“ gesammelten Erfahrungen zeigen, dass dieser Prozess nach drei bis fünf Durchläufen zum Abschluss kommt.

Ablauforganisatorische Aspekte

Die Durchführung der Geheimhaltungsprüfungen ist eine äußerst komplexe und arbeitsintensive Aufgabe, die sich an die Datenaufbereitungsphase anschließt. Das Spektrum der in der Praxis angewandten Vorgehensweisen reicht hierbei von der rein manuellen Durchführung der Geheim-

haltungsprüfungen bis zur komplett maschinellen Abwicklung. Eine systematische Einbettung der Geheimhaltungsprüfungen in einen automatisierten Workflow liegt in der Regel aber nicht vor. Dies galt auch für das in der Vergangenheit im Statistischen Landesamt Rheinland-Pfalz – auf Großrechnerplattform entwickelte und eingesetzte – Geheimhaltungsverfahren. Diese Feststellung verdeutlicht ein zentrales – neben der Migration auf die Client-Server-Plattform angestrebtes – Entwicklungsziel des neuen Geheimhaltungsverfahrens: „Einbettung der Geheimhaltungsprüfungen in einen weitgehend automatisierten Workflow bei gleichzeitiger Nutzung der bereits im LIS gespeicherten Meta-Daten“.

Info

In der nächsten Ausgabe der Statistischen Monatshefte wird der Workflow der LIS-Geheimhaltung vorgestellt.

Harald Wirtz leitet die Abteilung 3 „Wirtschaft, Landwirtschaft, Umwelt, Finanzen, Steuern“.

Claudia Baier ist Programmiererin im Sachgebiet „Informationssysteme“.