

Neues Geheimhaltungsverfahren des Statistischen Landesamtes

Teil 2: Der Workflow der LIS-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. Bisher wurde im Statistischen Landesamt bei der Durchführung der Geheimhaltungsprüfungen ein großrechnerbasiertes Verfahren eingesetzt. Aufgrund seiner mangelnden Integration in den Aufbereitungsprozess des Landesinformationssystems (LIS) und der im Verbund zu Ende gehenden Großrechner-Ära wurde im Jahr 2009 die Konzeption und Programmierung eines neuen Geheimhaltungsverfahrens in Angriff genommen.

Nachdem in der letzten Ausgabe der Statistischen Monatshefte die methodischen Hintergründe des in das Landesinformationssystem (LIS) integrierten Geheimhaltungsverfahrens dargestellt wurden, steht in diesem Beitrag der zu praktizierende Workflow im Vordergrund.

Das Verfahren zur Abwicklung des Workflows der LIS-Geheimhaltung umfasst fünf Anwendungsbereiche. Diese haben die Administration des Systems, die Meta-Datenverwaltung, die Auswertung des sogenannten Datenpools, die ETL-Datenübernahme und die Durchführung der Geheimhaltungsprüfungen i. e. S. zum Gegenstand.

Diese Funktionalitäten werden in einem geschlossenen Programmsystem und unter Verwendung einer vereinheitlichten Benutzungsoberfläche abgewickelt. Hierbei werden vielfach Programmmodule benutzt, die ursprünglich für das LIS-Auskunftssystem

entwickelt wurden. Diese Module wurden so ausgelegt, dass sie automatisch erkennen, ob sie im Rahmen des LIS-Auskunftssystems oder im Zuge der Geheimhaltungsprogramme zum Einsatz kommen. Unterschiede ergeben sich für den Anwender lediglich hinsichtlich der Benennung der Bildschirmmasken in der Kopfzeile, in der Farbe einzelner Menüpunkte sowie einzelner Zusatzfunktionen. Auf diese Weise ließen sich Doppelentwicklungen vermeiden. Das LIS-Geheimhaltungsverfahren basiert darüber hinaus auf dem gleichen Datenbankmodell wie das LIS-Auskunftssystem. Beide Anwendungen benutzen die gleichen

LIS-Geheimhaltung, integratives Programmsystem mit standardisierter Benutzungsoberfläche

G 1

Programme



Meta-Daten. Die Steuerung des Geheimhaltungsprogramms erfolgt ausschließlich menügestützt. Aufbau, Pflege und Zuführung externer Steuerungsdateien sind nicht erforderlich.

Die LIS-Geheimhaltung wird unter Zuhilfenahme eines separaten Rechnersystems und einer eigenständigen Datenbank abgewickelt.¹ Die durchzuführenden Prüfungen sind sehr ressourcenintensiv. Deshalb ist die parallele Abwicklung auf dem für Auskunftszwecke eingesetzten LIS-Produktionsserver nicht möglich. Zudem stellte sich während der Verfahrensentwicklung heraus, dass die Prozeduren der Geheimhaltungsprüfung eine andere Datenbankparametrisierung als die Routinen zur Abwicklung der Auskunftsdatenbank erfordern.

Die nachstehenden Erläuterungen beschreiben auszugsweise den Workflow des Geheimhaltungsverfahrens differenziert nach den einleitend genannten fünf Funktionsbereichen. Dies in dem Bewusstsein, dass die LIS-Geheimhaltung eine Erweiterung des in der Vergangenheit entwickelten Workflows zur Definition und Befüllung von Datenquardern des LIS ist. Zum Zwecke der Illustration wird auf das Beispiel „Umsatzsteuerstatistik der Voranmeldungen“ zurückgegriffen.

Durchführung der LIS-Geheimhaltung erfolgt auf separatem Server

Meta-Datenverwaltung = Aufbau eines Datenpools

Über die Einrichtung von „Eigentümern der Datenpools“ und explizite Freigaben wird – analog zum LIS – die Rechtevergabe zur Nutzung des LIS-Geheimhaltungsverfahrens gesteuert. Derartige Festlegungen sind allerdings nur dann notwendig, wenn die Geheimhaltungsarbeiten von verschiedenen Personen durchgeführt werden und die Mitarbeiterinnen und Mitarbeiter untereinander keinen Zugriff auf die Daten der anderen haben sollen. Wird diese Differenzierung nicht gewünscht, kann das Geheimhaltungsprogramm unter einer Benutzergruppe zur Anwendung kommen. Dieser sind dann alle Benutzer zugeordnet; sie sind zugleich Eigentümer aller Datenpools.

Definitionen

ETL = Extraktion, Transformation, Laden: Software im Sinne eines Bindeglieds zwischen der Datenbasis der operativen Ebene (z. B. der mittels Verbundprogrammen erzeugten Daten) und der informativen Datenbasis des LIS.

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.

¹ Zum Einsatz kommt ein Rechnersystem DELL PowerEdge R710 mit einem Quad-core-Prozessor Intel Xeon X5550, 16 GB Arbeitsspeicher und einer Plattenkapazität von einem Terabyte. Als Datenbanksystem wird ORACLE Version 11g verwendet.

G 2 Übersicht Datenquader

73311 Umsatzsteuer (Vorabmeldung)

Übersicht der Datenquader zur Erhebung

Name	Bezeichnung	geheim	Daten	ext. Name	Eigentümer	Zeitd. MM	W
73311DP401	Umsatzsteuerpflichtige und Umsatz nach WZ-2003, Größenklasse des Umsatz LuL und Rechtsform		a		GH-733	Jahr	
73311DP402	test Datenpool				GH-733	Jahr	
73311GJ020	Umsatzsteuer: 2 Hierarchien (Region, WZ) und 3 geheime WMM		a		GH-733	Jahr	
73311GJ030	Umsatzsteuer: 3 Hierarchien (Region, WZ, GKL)		a		GH-733	Jahr	
73311GJ040	Umsatzsteuer: 4 Hierarchien (Region, WZ, GKL, RP)		a		GH-733	Jahr	
73311GJ050	Umsatzsteuer: 5 Hierarchien (Region, WZ, GKL LuL, Rechtsform, GKL Umsatz insg) 3 gh Wmm		a		GH-733	Jahr	
73311HF030	Umsatzsteuer Hary's Testquader				GH-733	Jahr	
73311HF031	Umsatzsteuer Hary				GH-733	Jahr	

Pflege der Quaderdefinitionen

73311DP401 Name extern WZ-2003

Bezeichnung: Umsatzsteuerpflichtige und Umsatz nach WZ-2003, Größenklasse des Umsatz LuL und Rechtsform

2. Bezeichnung: Umsatzsteuerpflichtige und Umsatz nach WZ-2003, Größenklasse des Umsatz LuL und Rechtsform

Name: 73311DP401

Eigentümer: GGHaltung Umsatzsteuer

Datenebene: TB_GH_GE4 (max. 20 Dimensionen und 100 Wertmerkmale)

Zeitkriterium: Jahr

Anzahl Abrufe: 18
Letzter Abruf: 01.10.2010 10:32:19

07.01.2011 15:10:00 ADMIN Zuweisen von Merkmalen zu einem Datenquader.

G 3 Dimensionen

73311 Umsatzsteuer (Vorabmeldung)

Übersicht der Datenquader zur Erhebung

Name	Bezeichnung	geheim	Daten	ext. Name	Eigentümer	Zeitd. MM	W
73311DP401	Umsatzsteuerpflichtige und Umsatz nach WZ-2003, Größenklasse des Umsatz LuL und Rechtsform		a		GH-733	Jahr	
73311DP402	test Datenpool				GH-733	Jahr	
73311GJ020	Umsatzsteuer: 2 Hierarchien (Region, WZ) und 3 geheime WMM		a		GH-733	Jahr	
73311GJ030	Umsatzsteuer: 3 Hierarchien (Region, WZ, GKL)		a		GH-733	Jahr	
73311GJ040	Umsatzsteuer: 4 Hierarchien (Region, WZ, GKL, RP)		a		GH-733	Jahr	
73311GJ050	Umsatzsteuer: 5 Hierarchien (Region, WZ, GKL LuL, Rechtsform, GKL Umsatz insg) 3 gh Wmm		a		GH-733	Jahr	
73311HF030	Umsatzsteuer Hary's Testquader				GH-733	Jahr	
73311HF031	Umsatzsteuer Hary				GH-733	Jahr	

Pflege der Quaderdefinitionen

Dimensionen

nicht zugeordnete Erhebungs-Dimensionen

Fachschl.	Bezeichnung
wZ03 T4	wZ03 Klassen
FP-GM0	Gemeinde
LS-GK2	GKL Umsatz insgesamt
wZ03 T5	wZ03 Unterklassen
wZ03 T3	wZ03 Gruppen

Quader-Dimensionen

Nr.	Mer.St.	Fachschl.	Bezeichnung
1	1	LAND	Bundesland
2	1	RP SR	Statistische Region
3	1	RP RS	Kreis
4	1	RP VG	Verbandsgemeinde
5	2	wZ03 T0	wZ03 Abschnitt
6	2	wZ03 T1	wZ03 Unterabschnitt
7	2	wZ03 T2	wZ03 Abteilung
8	3	USTR01	GKL Umsatz LuL
9	4	USTR01	Rechtsform

07.01.2011 15:10:00 ADMIN Nicht zugeordnete Dimensionen.

**Merkmale
und deren Aus-
prägungen**

Vorbereitend zur Datenpoolerstellung werden die erforderlichen Merkmale (z. B. die Kreisebene) definiert und diesen Merkmalsausprägungen (z. B. die Bezeichnungen der Kreise) zugeordnet. Die Merkmalsausprägungen werden hierbei aus der LIS-Auskunftsdatenbank in die Geheimhaltungsdatenbank importiert. Dies ist für die Konsistenz und Effizienz der Meta-Datenpflege beider Datenbanksysteme von großer Bedeutung.

Die der Geheimhaltungsprüfung zu unterziehenden Daten einer Statistik müssen zunächst in einen Datenpool eingelagert werden. Dieser stellt gewisse strukturelle Anforderungen. Er wird mit dem Quaderdefinitionsprogramm des LIS erstellt, wobei zunächst eine allgemeine Beschreibung erfolgt. Sie umfasst z. B. Angaben zur Bezeichnung, zum Eigentümer und zur Zeitidentifikation.

**Datenpool, eine
strukturierte
Beschreibung
der unter
Geheimhal-
tungsaspekten
zu prüfenden
Daten**

Im nächsten Schritt sind die regionalen, sachlichen und zeitlichen Hierarchien zu definieren. Eine zwingende Voraussetzung ist, dass die hierdurch beschriebenen Datenklassifizierungen eine durchgängige Summierung erlauben. Dies ist z. B. bei der Hierarchie „Bundesland – Statistische Regionen – Kreisebene – Verbandsgemeindeebene – Ortsgemeindeebene“ der Fall. Auf der Grundlage dieser Definitionen werden in die Prüfungen zur Sicherstellung der sekundären Geheimhaltung alle Rand- und Zwischensummen eines n-dimensionalen Datenquaders einbezogen.

Hinsichtlich des Beispiels der Umsatzsteuerstatistik (Voranmeldungen) ergibt sich folgendes Bild:

- Hierarchie-1 bildet über vier Stufen (Bundesland, Statistische Regionen, Kreisebene und Verbandsgemeindeebene) die regionalen Beziehungen des Datenraums ab.

- Hierarchie-2 beschreibt über drei Stufen (WZ-Abschnitt, WZ-Unterabschnitt, WZ-Abteilung) die im Datenpool abgebildete Wirtschaftszweigsystematik.
- Hierarchie-3 ist einstufig und beinhaltet die Größenklassen.
- Hierarchie-4 ist eindimensional und enthält die Rechtsformen.

In der Terminologie der LIS-Auskunftsdatenbank deckt der Datenraum somit über vier Hierarchien neun Dimensionen² ab. Maßgeblich für die Gliederungstiefe mehrdimensionaler Hierarchien ist die gewünschte Granularität der Auswertungen.

Im Beispiel sollen bis auf Verbandsgemeindeebene alle regionalen Stufen veröffentlicht werden; sachlich wird die Abteilung als tiefste Gliederung der Wirtschaftszweige vorgesehen. Durch die Wahl der Reihenfolge der sachlichen Hierarchien lässt sich eine gewisse Wertigkeit der Klassifikationen festlegen. Der sekundäre Sperralgorithmus sperrt tendenziell eher auf der Hierarchie-4 als auf der Hierarchie-2.

Folgende Restriktionen sind zu beachten:

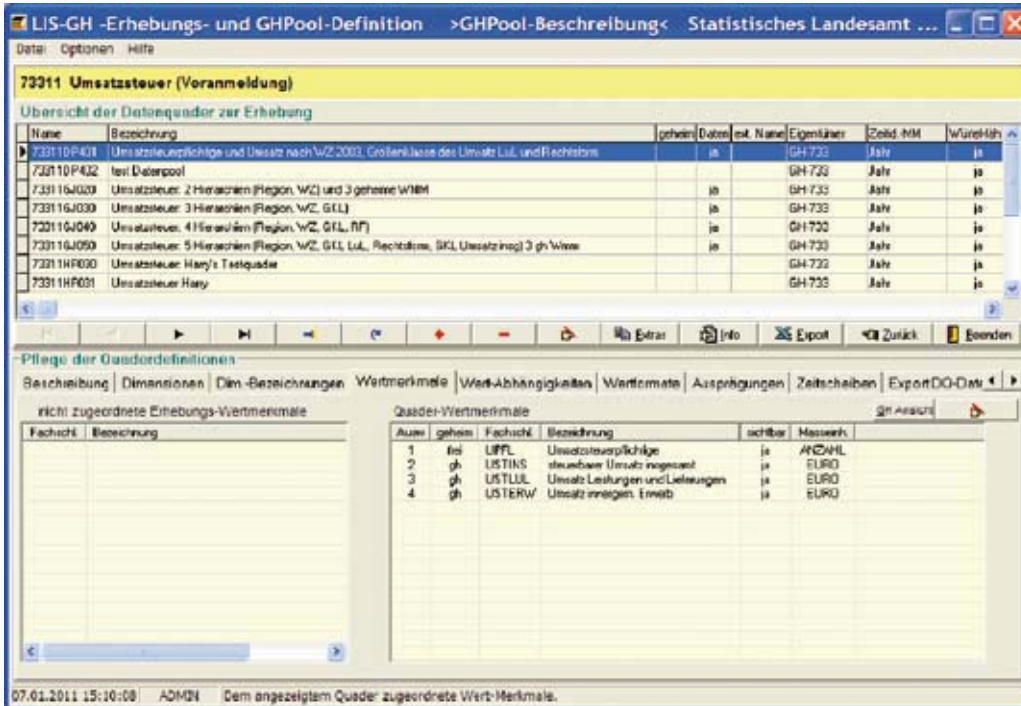
- Die regionale Gliederung muss stets die Hierarchie-1 bilden.
- Die oberste Ebene der regionalen Hierarchie muss immer die Verdichtung auf einen einzelnen Wert ermöglichen; darf also nur eine Ausprägung enthalten. Im Regelfall handelt es sich hierbei um die Landessumme, also das Bundesland.
- Die Zeitklassifikation (z. B. Monate bei unterjährigen Statistiken) muss – falls vorhanden – auf der zweiten Hierarchiestufe definiert werden.
- Nach der aktuellen Bedarfsanforderung sieht der gegenwärtige Ausbaustand des Verfahrens maximal sieben Hierarchien vor.

² In der LIS-Auskunftsdatenbank wird z. B. jede regionale Ebene (Bundesland, ..., Ortsgemeinde) als eine Dimension bezeichnet

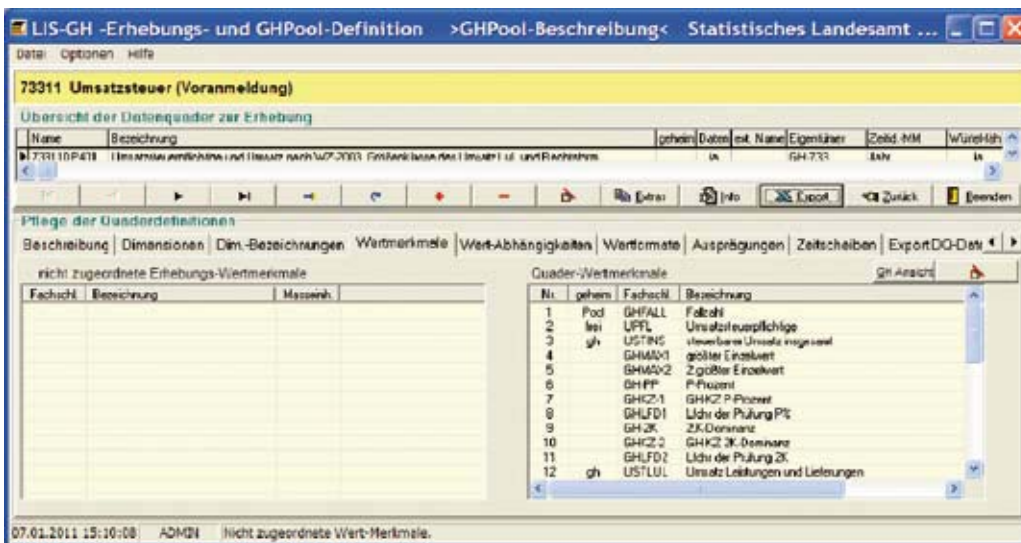
Dem Datenpool sind schließlich die Wertmerkmale des Datenraums zuzuordnen. Dabei wird zwischen Wertmerkmalen, die der Geheimhaltung unterliegen und freien Wertmerkmalen unterschieden. Im Bei-

spiel der Umsatzsteuerstatistik sind dies die Anzahl der Umsatzsteuerpflichtigen als freies Merkmal sowie der Umsatz insgesamt, der Umsatz aus Lieferungen und Leistungen und der Umsatz aus innergemeinschaft-

G 4 Wertmerkmale



G 5 Info-Merkmale



Ü 1

Datenpoolbeschreibung „Umsatzsteuerstatistik (Vor Anmeldungen)“

Lfd. Nummer	Inhalt
1	Fallzahl
2	Umsatzsteuerpflichtige (freier Wert)
3	Umsatz insgesamt
4 bis 11	Info-Wertmerkmale zum Umsatz insgesamt
12	Umsatz aus Lieferungen und Leistungen
13 bis 20	Info-Wertmerkmale zum Umsatz aus Lieferungen und Leistungen
21	Umsatz aus innergemeinschaftlichem Erwerb
22 bis 29	Info-Wertmerkmale zum Umsatz aus innergemeinschaftlichem Erwerb

lichem Erwerb als geheimhaltungskritische Merkmale.³ Nach Abschluss dieser Definition werden automatisch ein Wertmerkmal Fallzahl und zu jedem geheim zu haltenden Merkmal zusätzlich acht Info-Wertmerkmale angelegt. Für das Beispiel erhält der Datenpool somit die in der Übersicht gezeigte Struktur.

Berücksichtigung von Abhängigkeiten zwischen Wertmerkmalen

Abhängigkeiten zwischen Wertmerkmalen, die bei der sekundären Geheimhaltungsprüfung von Relevanz sind, werden über eine entsprechende Menüfunktion definiert. Dabei werden einem Basismerkmal ein oder mehrere abhängige Wertmerkmale zugeordnet. In der Beispielanwendung setzt sich der Umsatz insgesamt aus dem Umsatz aus Lieferungen und Leistungen sowie dem Umsatz aus innergemeinschaftlichem Erwerb zusammen. Im sekundären Geheimhaltungslauf wird aufgrund der vorgenommenen Definition dem Wert Umsatz aus Lieferungen und Leistungen eine höhere Wertigkeit als den beiden anderen Umsatzmerkmalen zugewiesen, d. h. bei notwendigen Sperrungen werden die beiden abhängigen Werte vorrangig von der Anzeige ausgenommen.

³ Diese Differenzierung wurde lediglich bezüglich der diesen Ausführungen zugrundeliegenden Testanwendung gewählt. Beim endgültigen Produktivlauf wurden alle vier Merkmale als geheimhaltungskritisch deklariert.

Sollen sich sachliche Hierarchiekombinationen nicht in den später zu befüllenden LIS-Auskunftsquadern bzw. sonstigen Tabellierungen wiederfinden, können diese Kombinationen gezielt von der Geheimhaltungsprüfung ausgenommen werden. So lassen sich gerade bei schwach besetzten sachlichen Klassifikationen unnötige Sperren vermeiden. In der Beispielanwendung soll die tiefste sachliche Gliederung nach Wirtschaftszweigen und Größenklassen, bzw. Wirtschaftszweigen und Rechtsformen erfolgen. Eine Kombination von Größenklassen und Rechtsformen ist in keinem Auskunftsquader vorgesehen.

Datenübernahme

Das in die Geheimhaltungsprüfung einzubeziehende Datenmaterial muss auf Einzelsatzebene vorliegen. Außerdem muss je Erhebungseinheit ein Identifikationskriterium – z. B. eine anonymisierte Betriebsnummer – im Datensatz existieren oder beim ETL-Vorimport über den Satzähler generiert werden.

Die Aufnahme der Daten in die Geheimhaltungsdatenbank erfolgt analog zum LIS über die ETL-Verfahrensmodule. Zu diesem Arbeitsgang gehören die Beschreibung der

LIS-GH -Erhebungs- und GHPool-Definition >GHPool-Beschreibung< Statistisches Landes...

Datfel Optionen Hilfe

73311 Umsatzsteuerstatistik (Vorankmeldungen)

Übersicht der Datenquader zur Erhebung

Name	Bezeichnung	geheim	Daten	est. Name	Eigentümer	Zeichn.-MM	Würfel-fähig	Info	Daten-tabelle
73311DP401	Umsatzsteuerpflichtige und Umsatz nach WZ-2003, Größenklasse des Umsa		ja		GH-733	Jahr	ja		TB_GH_QE4
73311DP402	test Datenpool				GH-733	Jahr	ja		TB_GH_QE4

Pflege der Quaderdefinitionen

Beschreibung | Dimensionen | Dim-Bez. | Wertmerkmale | GH-Wert-Abh. | Wertformate | Ausprägungen | Zeitscheiben | GH-Hierarchien | GH-Exp. < >

Wert-Abhängigkeiten

IdNr.	Basis	IdNr. Abh. Wert	Bezeichnung	Letzte Änderung	
12	USTLUU		Umsatz LuL	26.01.2011 15:10:30	
		3	USTINS	Umsatz insgesamt	26.01.2011 15:10:30
		21	USTERW	Umsatz innerErwerb	26.01.2011 15:10:30

07.02.2011 13:42:11 ADMIN Änderungsdienst der Quader.

LIS-GH -Erhebungs- und GHPool-Definition >GHPool-Beschreibung< Statistisches Landes...

Datfel Optionen Hilfe

73311 Umsatzsteuerstatistik (Vorankmeldungen)

Übersicht der Datenquader zur Erhebung

Name	Bezeichnung	geheim	Daten	est. Name	Eigentümer	Zeichn.-MM	Würfel-fähig	Info	Daten-tabelle
73311DP301	Umsatzsteuer für RLP nach WZ-2003 und Größenklasse Umsatz LuL (3 Hier		ja		GH-733	Jahr	ja		TB_GH_QE1
73311DP401	Umsatzsteuerpflichtige und Umsatz nach WZ-2003, Größenklasse des Umsa		ja		GH-733	Jahr	ja		TB_GH_QE4
73311DP402	test Datenpool				GH-733	Jahr	ja		TB_GH_QE4
73311GJ020	Umsatzsteuer: 2 Hierarchien (Region, WZ) und 3 geheime WMM		ja		GH-733	Jahr	ja		TB_GH_QE1
73311GJ040	Umsatzsteuer: 4 Hierarchien (Region, WZ, GKL, RF)		ja		GH-733	Jahr	ja		TB_GH_QE4

Pflege der Quaderdefinitionen

Beschreibung | Dimensionen | Dim-Bez. | Wertmerkmale | GH-Wert-Abh. | Wertformate | Ausprägungen | Zeitscheiben | GH-Hierarchien | GH-Exp. < >

Ausschluss Hierarchiestufen-Kombination aus GH-Prüfung

Kombi-Nr	Hierarchie	nicht mit Hier.	1 Merkmal der Hierarchie
1	3		GKL Umsatz LuL
1		4	Rechtsform

07.02.2011 09:43:39 ADMIN Exportieren nach MS-Excel.

G7 Regionalmerkmale, Ausprägungen

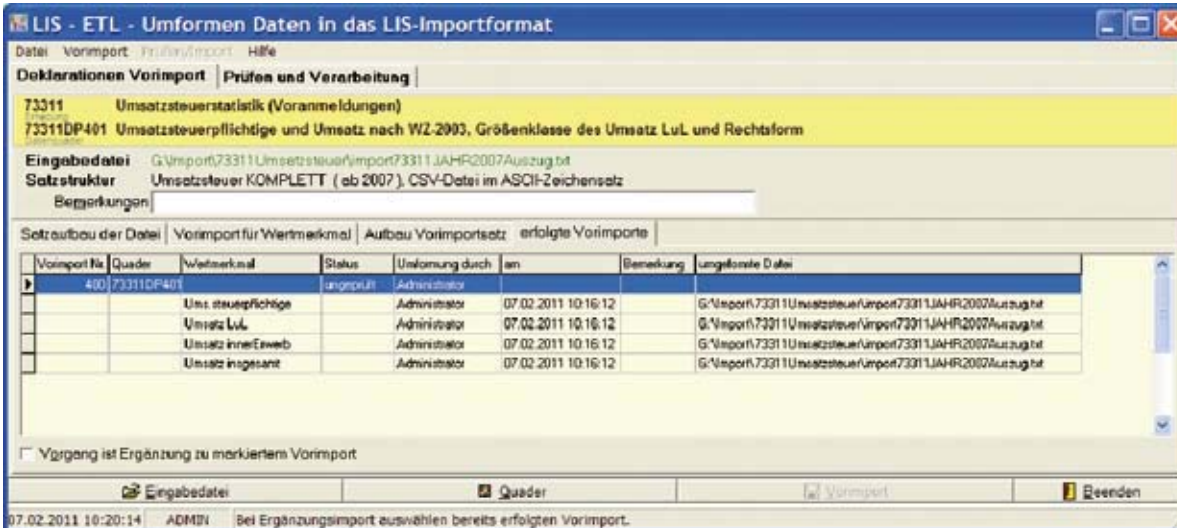
The first screenshot shows a list of regional features (Regionalklassifizierende Merkmale) with columns for Fachschlüssel, Kurzbezeichnung, Bezeichnung, Maßeinheit, Typ, externer Fachschl., and Info. The second and third screenshots show the 'Zuordnung und Struktur' (Assignment and Structure) dialog boxes for 'Rheinland-Pfalz' and 'Statistische Region' respectively, showing the mapping of specific features to these regional categories.

G8 Wertmerkmale

The screenshot displays the 'Wert-Merkmale' (Value Features) section. It includes a table of features with columns for Fachschlüssel, Kurzbezeichnung, Bezeichnung, Maßeinheit, Typ, externer Fachschl., and Info. Below the table, there are navigation and action buttons (Info, Nutzung, Export, Referenzen) and a detailed view for the selected feature 'GHFALL' (Fallzahl).

Fachschlüssel	Kurzbezeichnung	Bezeichnung	Maßeinheit	Typ	externer Fachschl.	Info	Letzte Änd.
GH-2K	2,K-Dominanz	2,K-Dominanz	Prozent				08.01.200
GH-PP	P-Prozent	P-Prozent	Prozent				08.01.200
GHFALL	Fallzahl	Fallzahl	Anzahl				17.06.200
GHKZ-1	GH-KZ P-Prozent	GH-KZ P-Prozent	Anzahl				08.01.200
GHKZ-2	GH-KZ 2K-Dominanz	GH-KZ 2K-Dominanz	Anzahl				08.01.200
GHLFD1	Prüf.Lfdnr P%	Lfdnr der Prüfung P%	Anzahl				19.08.200
GHLFD2	Prüf.Lfdnr 2K	Lfdnr der Prüfung 2K	Anzahl				19.08.200
GHMAX1	größter Einzelwert	größter Einzelwert	Anzahl				08.01.200

Selected Feature Details (GHFALL):
 Fachschlüssel: GHFALL, Kurzbezeichnung: Fallzahl, Bezeichnung: Fallzahl, Maßeinheit: Anzahl, Typ: Anzahl.



Eingabedateien, die Definition von Umformungsregeln⁴, die Verbindungsbeschreibung „Datensatz zu Datenpool“ und schließlich der ETL-Vorimport der Einzelsätze.

Nach erfolgreicher Einlagerung der Einzelsätze wird ein sogenannter Geheimhaltungsvorgang erstellt; auf seiner Grundlage werden die weiteren Schritte mittels des Moduls „Durchführung Geheimhaltung“ initiiert.

Durchführung der Geheimhaltung

Das Programm zeigt in der ersten Sicht alle Datenpools, zu denen es Geheimhaltungsvorgänge gibt. Nach Auswahl des gewünschten Datenpools werden auf der rechten Seite der Bildschirmanzeige alle dazugehörigen Vorgänge (z. B. die Geheimhaltungsprüfungen einzelner Statistikjahre betreffend) aufgelistet. Im Beispiel-Datenpool existiert ein Vorgang, der die Werte für das Jahr 2007 enthält.

Ein Doppelklick auf den Vorgang öffnet die weiteren Menüschritte zur Steuerung der

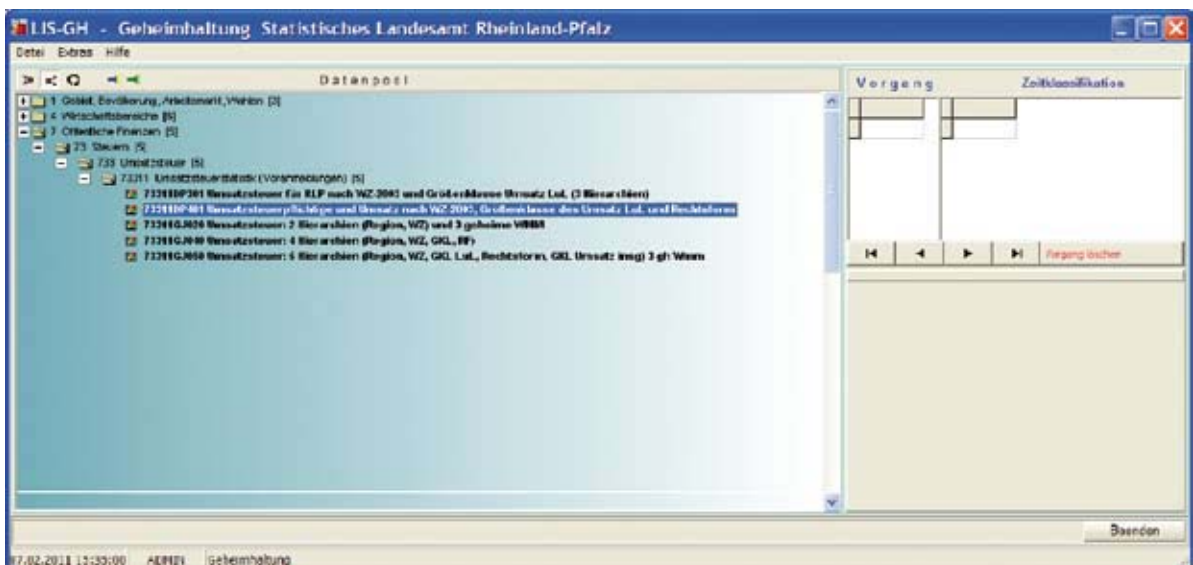
Geheimhaltungsprüfung. Zuerst muss die primäre Geheimhaltung initialisiert werden. Dazu gibt der Anwender die gewünschten Grenzwerte der primären Geheimhaltungsprüfung nach der p%- und der 2,k-Dominanzregel an. Die Primärsperren werden für beide Varianten in einem gemeinsamen Verarbeitungslauf gesetzt und gespeichert.

Bei der primären Geheimhaltungsprüfung erfolgt im ersten Schritt eine Generierung des kompletten Datenraums, der sich aus den Einzelsätzen und den über die Definition der Hierarchien resultierenden Summierungen ergibt. Für jedes auf Makroebene geheime Wertmerkmal werden über alle regionalen und sachlichen Ebenen/Summierungen hinweg die folgenden Informationswertmerkmale ermittelt und im Datenpool eingelagert:

- Anteil des größten Einzelberichts-pflichtigen,
- Anteil des zweitgrößten Einzelberichts-pflichtigen,
- p%-Wert und
- 2,k-Dominanzwert.

⁴ Importfunktionen erleichtern das Einrichten der Metadaten und synchronisieren die LIS-Auskunfts- und die LIS-Geheimhaltungsdatenbank.

G 10 Geheimhaltungsvorgänge und Vorgangstatus

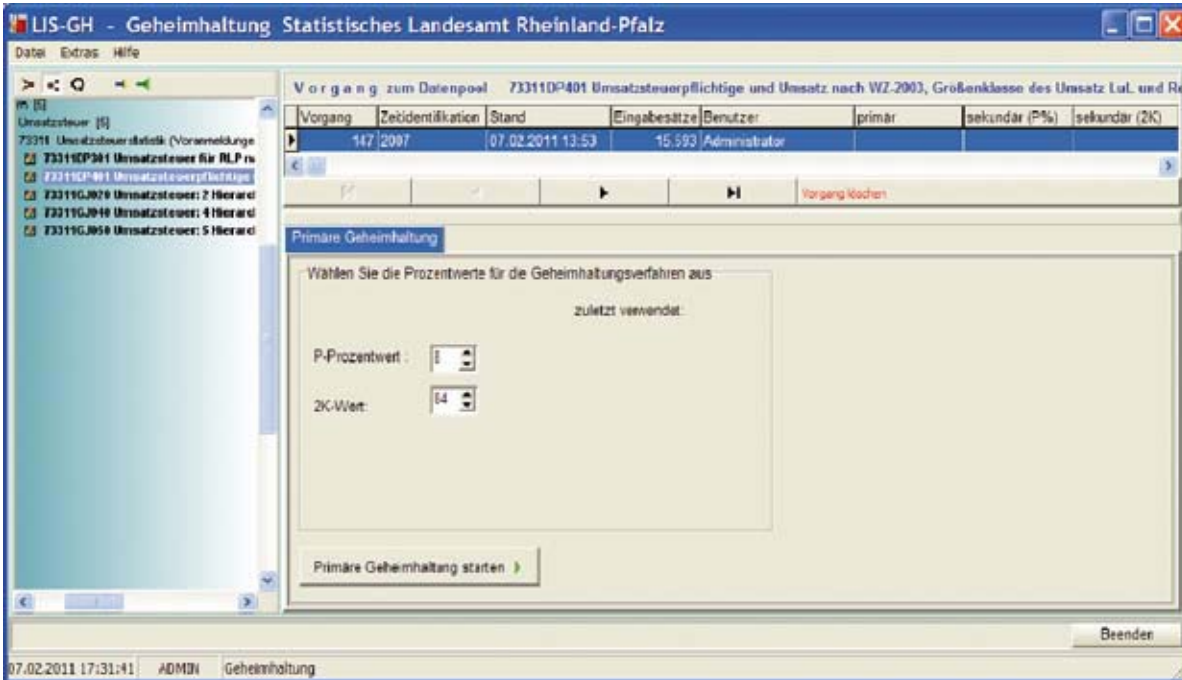


Anschließend werden die Geheimhaltungsfelder für die p%- und die 2,k-Dominanz anhand der vom Benutzer vorgegebenen Parameter bei Bedarf auf „primär geheim“ (Geheimhaltungskennzeichen 1 [GHZ 1]) gesetzt. Ein ausführliches Protokoll ist über die entsprechende Registerkarte abrufbar.

Vor dem Veranlassen der Sekundärprüfungen können eventuell vorhandene externe Sperren, wie sie z. B. aus einer bundeseinheitlich abgestimmten Geheimhaltung resultieren können, integriert werden.

Die Aufnahme der Sperren erfolgt analog zu dem Datenimport der Pooldaten über die ETL-Funktionen.

Im nächsten Verarbeitungsschritt wird die sekundäre Geheimhaltungsprüfung veranlasst. Hierbei ist zunächst vorzugeben, ob diese auf Basis der Ergebnisse der p%- oder der 2,k-Dominanzprüfung erfolgen soll. Zudem sind weitere Parameter zu setzen, die für die Wahl der sekundären Sperrpartner und die iterative Abwicklung der sekundären Geheimhaltungsprüfungen von Relevanz sind.



Vergabe differenzierter GHZ

Während des Geheimhaltungslaufs werden wiederum entsprechende Geheimhaltungskennzeichen vergeben:

- GHZ 2 = übernommene externe Sperre,
- GHZ 3 = Sperrung erster Durchlauf der sekundären Geheimhaltung,
- GHZ 4 = Sperrung zweiter Durchlauf der sekundären Geheimhaltung,
- GHZ n = Sperrung n-ter Durchlauf der sekundären Geheimhaltung⁵.

Alle Prüfungen einschließlich einer fortlaufenden Prüfungsnummer werden in einem Kurzprotokoll und in einem ausführlichen Protokoll festgehalten; ihre Anzeige wird durch Anklicken der entsprechenden Registerkarte veranlasst. Im Rahmen der Protokollierung wird u. a. die Anzahl der je Programmdurchlauf durchgeführten Sperren dokumentiert.

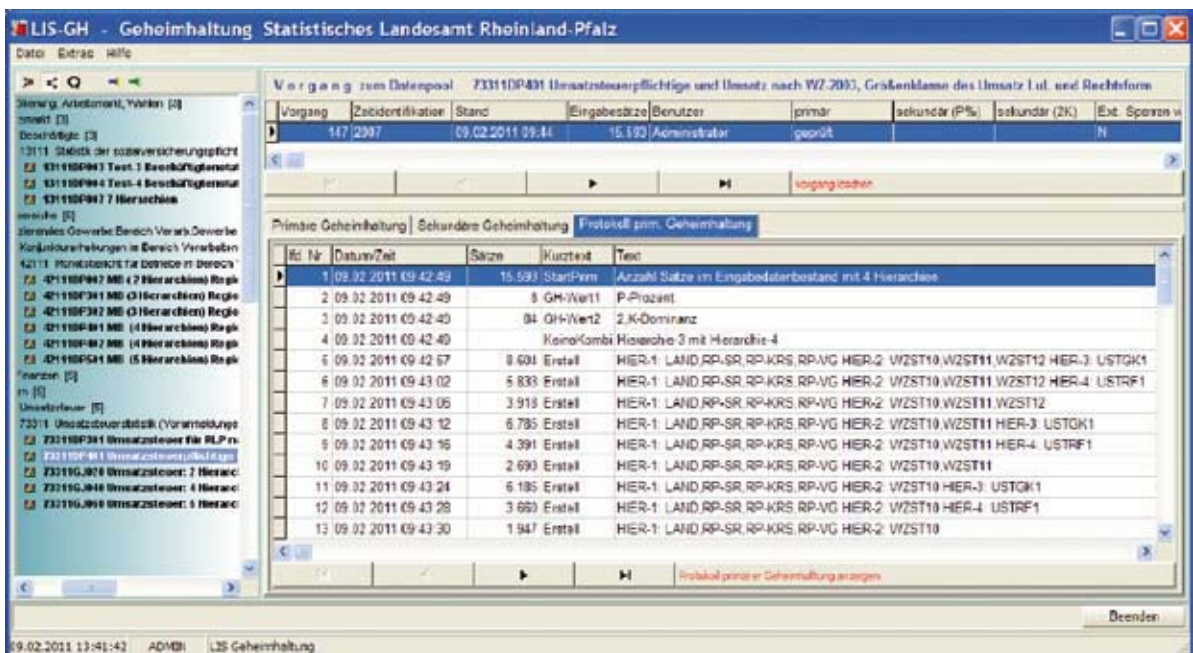
⁵ Die iterative Programmabwicklung endet, wenn keine weiteren Felder mehr sekundär zu sperren sind.

⁶ GENESIS = Gemeinsames NEues Statistisches Informationssystem.

Wiederholungsläufe der gesamten oder nur der sekundären Geheimhaltungsprüfungen mit veränderten Parametern zum Ausloten der erforderlichen Geheimhaltungstiefe sind möglich.

Nach Abschluss der sekundären Geheimhaltungsprüfung werden die geheimhaltungsqualifizierten Daten für die einzelnen zu befüllenden Auskunftsquader exportiert, um sie anschließend in die LIS-Auskunftsdatenbank über die Standard-Importfunktionen zu laden bzw. sonstigen Tabellierungsprogrammen zuzuführen. Die Überleitung der Daten erfolgt im GENESIS-Datenformat⁶. Hierbei ist folgendes zu beachten: Der Datenpool umfasst das komplette Datenspektrum des im Rahmen der Meta-Datenverwaltung definierten Datenraums, während in die LIS-Quader zu Auswertungs- und Veröffentlichungszwecken unter Umständen nur Ausschnitte der Dimensionen, Ausprägungen

G 12 Protokoll



G 13 Auswertung



und der Werte aufzunehmen sind. Im Beispiel der Umsatzsteuerstatistik wurden im Datenpool die sachlichen Dimensionen Wirtschaftszweigsystematik, Rechtsformen und Größenklassen eingelagert. Die Kombination „Rechtsformen/Größenklassen“ wurde im Rahmen der Datenpooldefinition

für die Geheimhaltungsprüfung aber gezielt „ausgesteuert“; beide Merkmale dürfen somit nicht in Kombination für einen einzelnen Quader exportiert werden. Die entsprechenden Exportbeschreibungen werden mittels des Datenpooldefinitionsprogramms festgelegt und gespeichert.

G 14 Abruf

Bundes...	Statistik...	Kreis...	WZ03 Abschlus...	Fakzahl/Anzahl	Umsatz insgesamt/Euro	GK+Z P-Prozent/Anzahl	Prof.Ums P%/Anzahl	Umsatz LUL/Euro	P-Prozent/Prozent 1 BK	GK+Z P-Prozent/Anzahl	Prof.Ums P%/Anzahl
007	Rheinland-Pf.	111 Koblenz, St.	A Landw. Forstwirtschaft	7		4	121		11,5	3	19
			D Verarbeit. Gewerbe	33					268,3		
			E Energie- Wasservers.	1		1			0,0	1	
			F Baugewerbe	20		3	7		67,4		
			G Handel, Informat. Rep.	118					251,7		
			H Gastgewerbe	32		5	7		205,5		
			I Veru. Nachr. überm.	29		5	7		206,0		
			J Rechts-Verw. gew.	1		1			0,0	1	
			K Grund- u. Wohn. wesen	142					216,2		
			L (Einkauf) Warenhandl.	5					84,5		
			M Ges. Ver- u. Soz. vers.	9					60,3		
			N Elektr. u. Fernkomm. Di.	21		3	7		72,7		
			«Sonstige»	438					808,7		
			A Landw. Forstwirtschaft	18		3	7		231,1		
			D Verarbeit. Gewerbe	58					214,2		
			E Energie- u. Wasservers.	1		1			0,0	1	

Auswertungen Datenpool

Transparenz der Geheimhaltung ist ein zentraler Vorteil des neuen Verfahrens

Nach Durchführung der primären und der sekundären Geheimhaltungsläufe liegen die der Prüfung unterzogenen und die hierbei zusätzlich generierten Daten (z. B. Geheimhaltungskennzeichen) im Datenpool auf Makroebene vor. Sie können mittels des Auswertungsmoduls des LIS zum Zweck der Qualitätssicherung zur Anzeige gebracht und analysiert werden. Diese Auswertungsmöglichkeit ist für die Transparenz der durchgeführten Geheimhaltungsprüfungen von zentraler Bedeutung. So kann z. B. kontrolliert werden, wie sich die Wahl alternativer p%- oder 2,k-Parameter konkret auf Merkmalsebene auswirkt. Ebenso wird das Setzen der Sekundärsperren transparent. Bei dem alten Geheimhaltungsverfahren des LIS war dies nicht gegeben. Konkrete Fragen

nach den Gründen bestimmter Sperrungen konnten häufig nicht oder nur mit hohem Aufwand beantwortet werden.

Weiteres Vorgehen

Das neue LIS-Geheimhaltungsverfahren wird derzeit einer Qualitätssicherung unterzogen. Danach werden die bislang mit dem alten Geheimhaltungsverfahren überprüften Statistiken sukzessive der neuen Anwendung zugeführt.

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

Claudia Baier ist Programmiererin im Sachgebiet „Informationssysteme“.