

Dipl.-Mathematiker Hartmut Rahm, Dipl.-Volkswirt Christian Zipse

Konzeption und Entwicklung von GENESIS

Das gemeinsam mit den Statistischen Ämtern der Länder entwickelte Informationssystem GENESIS (Gemeinsames Neues Statistisches Informationssystem des Bundes und der Länder) ist seit Anfang 2004 im Statistischen Bundesamt die zentrale Datenbank für die Bereitstellung und die Nutzung statistischer Ergebnisse. Darauf aufbauend konnte mit GENESIS-Online eine Internet-Auskunftsdatenbank für externe Nutzer eingeführt werden, die Tabellen als zentrales Mittel der Information auf Abruf erstellt, Dateien als Datenquader in weiterverarbeitbarer Form anbietet und den Zeitreihenservice des Vorgängersystems STATIS-BUND ersetzt.

Dieser Beitrag stellt das Basissystem, die Produktionsdatenbank, vor und geht auf seine Entwicklung, die damit verfolgten Ziele und seine heutigen Nutzungsmöglichkeiten ein. In einem weiteren Artikel in dieser Zeitschrift wird GENESIS-Online ausführlich beschrieben.

1 Qualitätsmerkmale eines Informationssystems

Bei der Entwicklung eines Informationssystems sind fachliche und organisatorische Vorgaben umzusetzen und technische Rahmenbedingungen einzuhalten.

Bezüglich der Entwicklung eines gemeinsamen statistischen Informationssystems der Statistischen Ämter des Bundes und der Länder sind dies die Einbindung in die Produktionsprozesse der amtlichen Statistik, die dezentrale Datenhaltung für regional differenzierte Daten und – in Anbetracht der Langfristigkeit der Datenbereitstellung

und der Vielfältigkeit der Rechnerlandschaften in den Statistischen Landesämtern und im Statistischen Bundesamt – auch die weitgehende Unabhängigkeit von Betriebssystemen und Rechnern eines Herstellers.

Bei der inhaltlichen Konzeption eines statistischen Informationssystems sind die Qualitätsmerkmale, die für alle statistischen Produkte der Statistischen Ämter des Bundes und der Länder gelten, eine bindende Leitlinie:

- 1. Relevanz:**
Die inhaltliche Auswahl und die regionale Differenzierung des Datenangebotes werden zwischen den Fachreferenten in Bund und Ländern abgestimmt; damit soll sichergestellt werden, dass sowohl die Möglichkeiten für interne Auswertungen als auch für Auswertungen externer Nutzer statistikspezifisch optimiert werden.
- 2. Genauigkeit:**
Die statistischen Daten werden grundsätzlich gemeinsam mit Qualitätskennzeichen, Definitionen und Erläuterungen angeboten, um die statistischen Aussagen korrekt interpretieren zu können.
- 3. Aktualität und Pünktlichkeit:**
Die Daten werden zu vorgesehenen Veröffentlichungsterminen termingerecht freigeschaltet; gleichzeitig werden die Daten für die automatisierte Erstellung von Tabellen im Rahmen der Pressemitteilungen genutzt.
- 4. Verfügbarkeit und Transparenz:**
Alle Daten, die keiner Geheimhaltung unterliegen, sind öffentlich zugänglich und mit allen notwendigen Erläuterungen versehen.

5. Vergleichbarkeit:

Bundes-, Länder-, Kreis- und Gemeindeergebnisse sind durch ein einheitliches Metadatenystem aufeinander abgestimmt. Das Datenbanksystem ist in jedem Statistischen Landesamt einsetzbar und garantiert vergleichbare Ergebnisse.

6. Kohärenz:

Die Verwendung einheitlicher Definitionen und Klassifikationen führt zu fachlich abgestimmten Sachverhalten und statistikübergreifender Konsistenz.

Die Standardisierung des Datenangebots über einheitliche Metadaten ist im Hinblick auf die zunehmende Vernetzung der Informationssysteme besonders wichtig. Bei der Nutzung stehen das schnelle Auffinden der Information, die gezielte Auswahl und die hohe Aktualität der Inhalte im Vordergrund; dies wird durch eine strukturierte Datenvorhaltung und vor allem dadurch gesichert, dass die Nachfrage nach Informationen nicht durch die Suche nach einer vorgefertigten Ergebnistabelle in einer Sammlung von Vortabellen bedient wird, sondern bei jeder Abfrage die gewünschte Information dynamisch aus der zentralen Datenbank erzeugt wird.

Für die Weiterverarbeitung der Tabellen ist aus Sicht der Nutzer die medienbruchfreie Übertragung der Tabellen in die gängigen PC-Formate besonders hilfreich.

2 Im Rückblick: STATIS-BUND

Mit einem Erlass des Bundesministeriums des Inneren von 1969 wurde der Aufbau eines Informationssystems im Statistischen Bundesamt angeordnet. Kernaufgabe von STATIS-BUND (Statistisches Informationssystem des Bundes) war die übersichtliche und leicht zugängliche Vorhaltung von Daten zu Auswertungszwecken.¹⁾

Die Datenhaltung erfolgte in der Regel in Form von einfach zu aktualisierenden und für Konjunkturbeobachtungen zweckmäßigen Zeitreihen mit umfassender Datendokumentation und einem separaten Definitionskatalog. 1976/77 war das System einsatzfähig und stand über Terminals mehreren hundert Mitarbeiterinnen und Mitarbeitern im Statistischen Bundesamt zur Verfügung. Daneben wurden im Laufe des über 25-jährigen Wirkbetriebs nach und nach Werkzeuge zur Analyse mit gängigen mathematisch-statistischen Verfahren zur Erzeugung statistischer Graphiken und zur Druckausgabe von Tabellen und Datenausdrucken entwickelt.²⁾

Die Verbreitungswege für datenbankgestützte Informationen haben sich seither grundlegend verändert; immer seltener werden solche Informationen in gedruckter Form zur Verfügung gestellt. Datenverbreitung findet überwiegend über das Internet statt. Das Statistische Bundesamt hat an dieser Entwicklung mit der Bereitstellung von tabellenorientierten Konjunkturdaten in Bildschirmtext (Btx) im Jahr 1993 sowie mit der Freischaltung der ersten Website im Jahr 1996 sehr

früh partizipiert.³⁾ Schon ein Jahr später wurde mit einem Bestellsystem ein Internetzugang zu den Zeitreihen in STATIS-BUND geschaffen, die über eine Mailbox zeitverzögert auf den eigenen PC geladen werden konnten.

Die Bereitstellung umfangreicher Datenbestände wurde durch die periodische Herausgabe fachlich differenzierter STATIS-BUND-CDs mit spezifischer Abfragesoftware sichergestellt. Am Ende des Betriebs von STATIS-BUND waren rund 180 000 Zeitreihen mit nahezu 100 Mill. Werten gespeichert. Diese Daten werden mit einer „historischen“ Ausgabe der STATIS-BUND-CD auf Dauer verfügbar bleiben.

3 Das Projekt GENESIS

3.1 Entwicklung

Im Zuge des Aufbaus der statistischen Landesämter in den neuen Bundesländern wurde das Ziel festgelegt, ein neues, für Bund und Länder einheitliches Informationssystem zu schaffen, weil die zu diesem Zeitpunkt eingesetzten, unterschiedlich strukturierten Systeme nicht mehr allen Anforderungen gerecht werden konnten. Im Beschluss zur Entwicklung eines Gemeinsamen Neuen Statistischen Informationssystems (GENESIS) wurde vorgegeben, dass die zentrale standardisierte Datenbank primär der fachübergreifenden Recherche, der Auskunftserteilung und der Bereitstellung von Daten zur Weiterverarbeitung dienen sollte.

Die Konzeption von GENESIS wurde von einer Arbeitsgruppe aus sechs Statistischen Landesämtern und dem für das Projektmanagement zuständigen Entwicklungsbüro im Statistischen Bundesamt erstellt. Nach der Vorstellung des Fachkonzepts wurden die grundlegenden Programmierarbeiten an einen externen Auftragnehmer vergeben. Die Anpassung der Software, zum Beispiel an die sehr hohen Speichervolumina in der amtlichen Statistik und die statistikspezifischen Variationen in der Tabellengestaltung, wurde arbeitsteilig in den Statistischen Landesämtern Baden-Württemberg, Bayern, Brandenburg und Sachsen und im Statistischen Bundesamt durchgeführt. Das Entwicklungsbüro koordinierte die Verbundtätigkeiten und führte die Programmmodule zusammen.

Nachdem 1997 die erste Version von GENESIS (Version 1.0) an alle interessierten statistischen Ämter ausgeliefert und in Betrieb genommen wurde, nahm eine Nutzergruppe aus Erstanwendern in den statistischen Ämtern ihre Arbeit auf. Deren Vorschläge gehen seitdem in die Weiterentwicklung jeder neuen Version ein und tragen zur Verbesserung von Funktionalität und Benutzerführung bei.⁴⁾

Die erste Version der Datenbanksoftware war noch eine zeichenorientierte Großrechneranwendung. Die neuen Anforderungen nach einer menügesteuerten, graphischen Nutzerführung und einer internetfähigen Datenbereitstellung

1) Siehe Füll, M./Kühn, J.: „Statistisches Informationssystem des Bundes – Neue Möglichkeiten zur schnelleren und rationelleren Informationsbereitstellung“ in ABI-Technik, Jahrgang 4 (1984), Nr. 3, S. 213 ff.

2) Siehe Schrey, E.: „Weiterentwicklung des Statistischen Informationssystems zu einem verteilten System aus APC und Großrechner“ in WiSta 12/1993, S. 888 ff.

3) Siehe Rahm, H.: „Statistik im Internet“ in WiSta 8/1996, S. 475 ff.

4) Siehe dazu Engelhardt, K./Langer, I.: „GENESIS ante portas“ in Bayern in Zahlen 1/2002, S. 25 ff.

fürten zu umfangreichen hausinternen Programmierarbeiten. Im Februar 2002 wurde mit der Version 1.6 eine Client-Server-Ausführung mit graphischer Benutzeroberfläche für den Testbetrieb ausgegeben. Im März 2002 folgte auf der CeBit-Messe in Hannover die Vorstellung der ersten Version von GENESIS-Online für den öffentlichen Zugang über das Internet. Schließlich wurde im Mai 2002 mit der Version 2.0 ein ausgereiftes Datenbankprogramm im Paket mit Client- und Online-Ausführung freigegeben (die Freischaltung im Internet erfolgte im Juli) und das Projekt abgeschlossen. Ein Pflege- und Wartungskonzept garantiert die weitere Verbesserung und die Erweiterung des Systems.

Die für einen reibungslosen Übergang auf GENESIS-Online benötigten Datenbankinhalte wurden bis Ende des Jahres 2003 konzipiert und eingespeichert. Damit löste das neue System den Datenhaltungsteil von STATIS-BUND, das zum 30. Juni 2004 stillgelegt werden konnte, endgültig ab. Die Daten sind archiviert und bleiben auf CD-ROM verfügbar.

3.2 Konzeption und Ziele

Das Ziel dieses Großprojekts im statistischen Verbund war in erster Linie eine einheitliche Datenbank mit gemeinsamen Inhalten und abgestimmten Metadaten. Dieses Ziel mag vielleicht simpel erscheinen, bedarf aber enormer Anstrengungen und vieler Kompromisse und führt in der konsequenten Umsetzung langfristig zu Qualitätsverbesserungen und zu einer verstärkten Kundenorientierung, denn die Datenbestände sind harmonisiert und die Tabellenerstellung führt auch bei dezentraler Datenhaltung zu einheitlich strukturierten Ergebnissen.

Statistikübergreifend abgestimmte Metadaten ermöglichen die Verknüpfung von Daten unterschiedlicher Erhebungen innerhalb des Datenbanksystems und bilden somit die Grundlage für weitergehende Analysen.

Das Metadatenystem bildet auch die Grundlage für eine Recherche nach Begriffen, Erläuterungen, Definitionen und wird durch Synonyme und Verweisungen in einem Thesaurus unterstützt.

Im Gegensatz zu STATIS-BUND, bei dem nach Möglichkeit alle anfallenden Aufgaben, wie zum Beispiel die Saisonbereinigung, innerhalb des Datenbanksystems gelöst wurden, unterstützt GENESIS mittels spezifischer Datenexportformate die Nutzung kommerzieller Auswertungsprogramme, mit denen die Kunden in der Regel bereits vertraut sind.

4 Die Speicherstrukturen der Datenbank

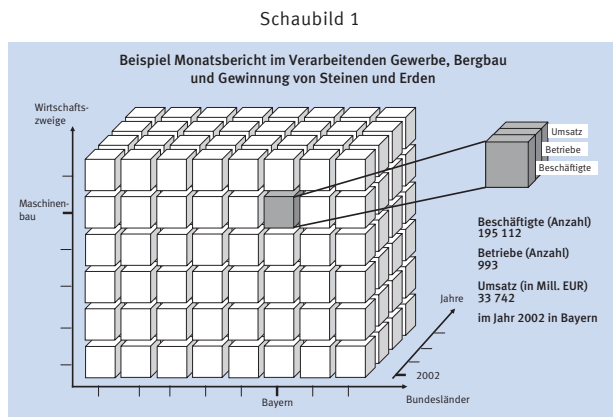
GENESIS basiert auf dem Datenbanksystem ADABAS der Software AG Darmstadt. ADABAS ist für den Betrieb auf Großrechnern unter diversen Betriebssystemen ebenso ausgelegt wie für UNIX-Systeme und damit grundsätzlich universell einsetzbar. Für die plattformübergreifende Kommunikation sorgt das Programm EntireX, eine so genannte Middleware aus dem gleichen Haus.

4.1 Datenmodell

Das grundlegende Modell bildet die Daten in einer mehrdimensionalen Speicherstruktur ab. Die von der amtlichen Statistik erhobenen Tatbestände weisen verschiedene Eigenschaften auf. Der eigentliche statistische Wert lässt sich abgrenzen nach sachlichen, räumlichen und zeitlichen Kriterien.

Im Monatsbericht für Betriebe im Bereich Verarbeitendes Gewerbe, Bergbau und Gewinnung von Steinen und Erden zum Beispiel lassen sich die gezählten Betriebe sachlich nach ihrer Zugehörigkeit zu einem Wirtschaftszweig und räumlich nach der Region ihres Standortes gliedern. Die Erhebungsmonate und Erhebungsjahre liefern die zeitliche Gliederung. Die erhobenen Wertmerkmale weisen also mehrere Dimensionen auf, deren Speicherstruktur am anschaulichsten in der dreidimensionalen Form eines Quaders abgebildet werden kann.

Ein Beispiel aus dem erwähnten Monatsbericht ist in Schaubild 1 dargestellt. Die Elemente des Quaders werden nach drei Achsen gegliedert, den Wirtschaftszweigen, der Regionalgliederung der Bundesländer und den gespeicherten Jahren. In jedem Quaderelement können mehrere Wertmerkmale abgelegt werden, wenn sie eine identische Erhebungsstruktur aufweisen. Die einzelnen Achsenabschnitte entsprechen den Ausprägungen, die ein Merkmal annehmen kann. Die im Beispiel dargestellte Regionalgliederung Bundesländer weist 16 Ausprägungen auf. Als Einzelkombination der Achsen Schnittpunkte können dann die gewünschten Werte gelesen werden. So gab es im Jahr 2002 in Bayern 993 Betriebe im Bereich Maschinenbau.



In der GENESIS-Datenbank können die „Quader“ theoretisch bis zu 20 Achsen und bis zu 20 Wertmerkmale aufweisen. Jede gespeicherte Statistik kann eine unterschiedliche Anzahl Quader verschiedener Größen und Kantenlängen besitzen. Mit jedem Wert werden die dazugehörigen Qualitätsanzeiger abgelegt. So ist genau definiert, ob es sich zum Beispiel um vorläufige, geheim zu haltende oder nicht verfügbare Daten handelt. Die Importdatensätze eines solchen Quaders können auch als Tabelle interpretiert werden, deren Gliederung ausschließlich aus Vorspalten besteht. In Tabelle 1 findet sich ein Beispiel für eine Importstruktur aus der Bevölkerungsstatistik.

Tabelle 1: Beispiel für die Importstruktur eines Quaders aus der Bevölkerungsstatistik

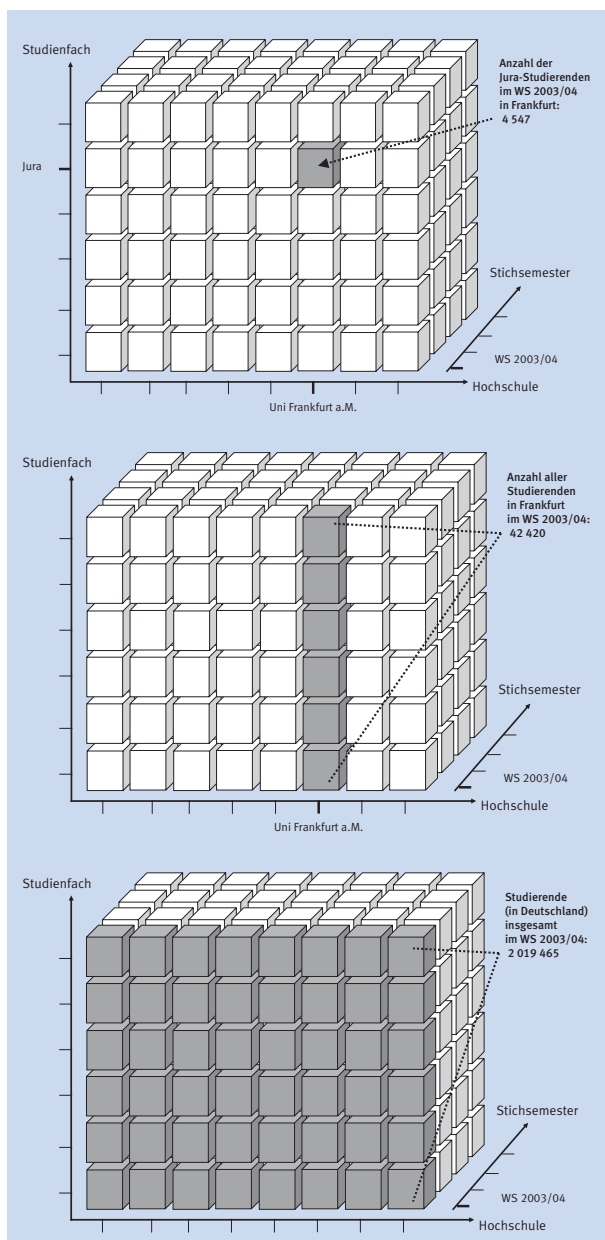
Bundesland	Altersjahre	Geschlecht	Stichtag	Anzahl Bevölkerung	Qualität	Sperrung
Hessen	36	Männlich	31.12.2003	54 543	endgültig	Nein
Hessen	36	Weiblich	31.12.2003	52 271	endgültig	Nein
Hessen	37	Männlich	31.12.2003	56 231	endgültig	Nein
Hessen	37	Weiblich	31.12.2003	53 386	endgültig	Nein
Hessen	38	Männlich	31.12.2003	56 508	endgültig	Nein
Hessen	38	Weiblich	31.12.2003	53 564	endgültig	Nein

4.2 Aggregation, Ableitung und Zugriff

Das Datenmodell ermöglicht eine zielgerichtete Auswertung der Daten ohne die redundante Speicherung von Rand- und Zwischensummen. In Schaubild 2 wird anhand eines Beispiels aus der Hochschulstatistik die Nutzung eines Quaders dargestellt. Je nach Fragestellung kann die gespei-

cherte Anzahl der Jura Studierenden an der Universität Frankfurt am Main im Wintersemester 2003/04 abgerufen werden, oder die über die Studienfächer aggregierte Anzahl aller Studierenden in Frankfurt am Main in diesem Semester. Über das so genannte „Weglassen von Achsen“ kann das Material beliebig verdichtet werden, im Beispiel bis zur Gesamtzahl der Studierenden in Deutschland im Wintersemester 2003/04. Diese Vorgehensweise wird in der Literatur auch *roll-up* oder *drill-up* genannt, das „Nach-oben-Wandern“ von den detaillierten Werten zu den Eckzahlen.

Schaubild 2



Bei Quadern ist es auch möglich, Darstellungen innerhalb festgelegter Achsenabschnitte zu bilden. In dem Quader zur Hochschulstatistik könnte die Entwicklung der Studierendenzahl im Fach Jura über die Zeit betrachtet werden, und dies für Frankfurt am Main oder jede andere gewünschte Universität oder auch für mehrere Universitäten oder Studienfächer. Man schneidet sozusagen „Scheiben“ aus dem Quader (*slicing*) oder entnimmt kleinere Quaderabschnitte (*dicing*) und bildet daraus Tabellen. Hierfür stellt GENESIS einen Tabellenassistenten zur Verfügung.

Eine weitere Bearbeitungsmöglichkeit bietet die Funktion der Regeln. Damit können individuelle Vorschriften geschaffen werden, nach denen ein Merkmal aus anderen Merkmalen abgeleitet werden kann. Es gibt zwei Typen von Regeln: den Algorithmus und die Referenz.

Die *Algorithmusregel* bildet aus Wertmerkmalen, wie zum Beispiel den Erwerbspersonen und der Bevölkerung, das neue Merkmal der Erwerbsquote oder es entsteht aus der Zahl von Ankünften und Übernachtungen in Hotels die durchschnittliche Aufenthaltsdauer.

Bei der *Referenzregel* werden die Ausprägungen eines klassifizierenden Merkmals einem Zielmerkmal neu zugeordnet. Die 16 Bundesländer könnten damit einerseits der Gliederung Früheres Bundesgebiet/Neue Länder oder andererseits der Aufteilung Stadtstaaten/Flächenstaaten zugeordnet werden. Eine Liste mit Altersjahren kann zu unterschiedlichen Altersklassen zusammengefasst werden. Mit Hilfe von gespeicherten Regeln werden je nach Fragestellung unterschiedliche Auswertungen aus dem gleichen unveränderten Basismaterial vorgenommen.

Durch die Eigenentwicklung der Datenbank konnten Besonderheiten der jeweiligen Statistik berücksichtigt werden. So unterscheidet GENESIS zum Beispiel zwischen Bestands- und Stromgrößen, da erstere nicht beliebig aggregierbar sind. Der Versuch eines Nutzers, aus dem monatlichen Bevölkerungsstand durch Summierung ein Jahresergebnis zu erzeugen, wird vom System abgewiesen.

Ein sehr differenzierter Zugangs- und Zugriffsschutz in sämtlichen Bereichen der Datenbank berücksichtigt die Anforderungen des Datenschutzes und der statistischen Geheimhaltung bei allen Anfragen interner und externer Stellen.

5 Die Funktionalitäten

Bei der Entwicklung von GENESIS wurden Funktionen eines Management-Informationssystems umgesetzt, die mit dem Begriff „Data Warehouse“ in den 1990er-Jahren bekannt geworden sind.⁵⁾ Das Konzept sieht einen zentralen Informationsserver vor, der fertig aufbereitete und strukturiert eingespeicherte Daten verwaltet. Mit Hilfe eines Metadatenreferenzsystems können gezielt und zeitnah gewünschte Informationen recherchiert und ausgewählt werden. Entscheidend ist der Zugriff über ein Client-Server-System, das die verteilte Nutzung aller Datenbestände erlaubt, das heißt die Umsetzung als übergreifendes Auskunftssystem für Daten und Metadaten, in dem wie in einem Warenhaus Informationen nach Bedarf zusammengetragen werden.

5.1 Tabellierung

Über das nach dem Einheitlichen Verzeichnis der amtlichen Statistiken (EVAS) strukturierte Verzeichnis der Erhebungen (siehe Schaubild 3) gelangt man zum Kernstück für die Auswertung eines Quaders, dem Tabellenassistenten. Dieser nutzt die gespeicherten Metadateninformationen zur Erstellung einer technisch optimierten Tabellenstruktur. Der Nutzer kann diese Tabellenstruktur seinen Auswertungsanforderungen individuell anpassen und die für ihn relevanten Ausprägungen auswählen. In Schaubild 4 wird diese Auswahl an einem Beispiel aus der Hochschulstatistik gezeigt. Nach dem Aufruf des Assistenten wird zunächst eine Vorauswahl der Achsen durchgeführt. Diese Ordnungsstruktur kann nun innerhalb der plausiblen Kombinationsmöglichkeiten des Datenmaterials nach Belieben verändert werden und so die gewünschte Tabelle spezifiziert werden. In Tabelle 2 sind zwei verschiedene Tabellenrahmen zu sehen, die aus dem gleichen Zahlenmaterial unterschiedliche Auswertungen generieren. In anderen Anwendungen sind solche Funktionen auch als Pivot-Tabellierung oder *pivoting* bekannt.

Schaubild 3

The screenshot shows the 'Genesis - Erhebungen' application window. It features a menu bar with options like 'Datei', 'Bearbeiten', 'Objekt', 'Sicht', 'Verzeichnis', 'Tabelle', 'Funktion', 'Recherche', and 'Hilfe'. Below the menu is a toolbar with various icons. The main area is titled 'Erhebungen' and contains 'Sichtoptionen' (View Options) with fields for 'Sortierkriterium' (EVAS), 'Auswahl', 'Sortierung' (auf), 'Verknüpfung mit', 'Kurztext', 'Rechte' (alle), and 'Typ' (alle). There are buttons for 'Zeigen', 'Optionen...', and 'Schließen'. The main display is a table with three columns: 'EVAS', 'Kurztext', and 'Datenquader'. The table lists various categories and their corresponding data counts. At the bottom, there is a status bar that reads 'Doppelklick:Eigenschaften, rechte Maustaste: Kontextmenü'.

EVAS	Kurztext	Datenquader
1	Gebiet, Bevölkerung, Erwerbstätigkeit, Wahlen	113
2	Bildung, Sozialleistungen, Gesundheit, Recht	122
21	Bildung und Kultur	12
211	Allgemeinbildende und berufliche Schulen	3
212	Berufsbildung	
213	Hochschulen	9
21311	Statistik der Studenten	7
21321	Statistik der Prüfungen	2
214	Ausbildungsförderung	
215	Weiterbildung	
216	Kultur	
219	Sonstiges im Bereich Bildung und Kultur	
22	Öffentliche Sozialleistungen	93
23	Gesundheitswesen	11
24	Rechtspflege	6
3	Wohnen, Umwelt	11
4	Wirtschaftsbereiche	201
5	Außenhandel, Unternehmen, Handwerk	112
6	Preise, Verdienste, Einkommen und Verbrauch	259
7	Öffentliche Finanzen	40
8	Gesamtrechnungen	54
9	Sonderbereiche	8

⁵⁾ Siehe z. B. Definition in Klußmann, N.: „Lexikon der Kommunikations- und Informationstechnik – Telekommunikation, Datenkommunikation, Multimedia, Internet“, Heidelberg 2000.

Schaubild 4

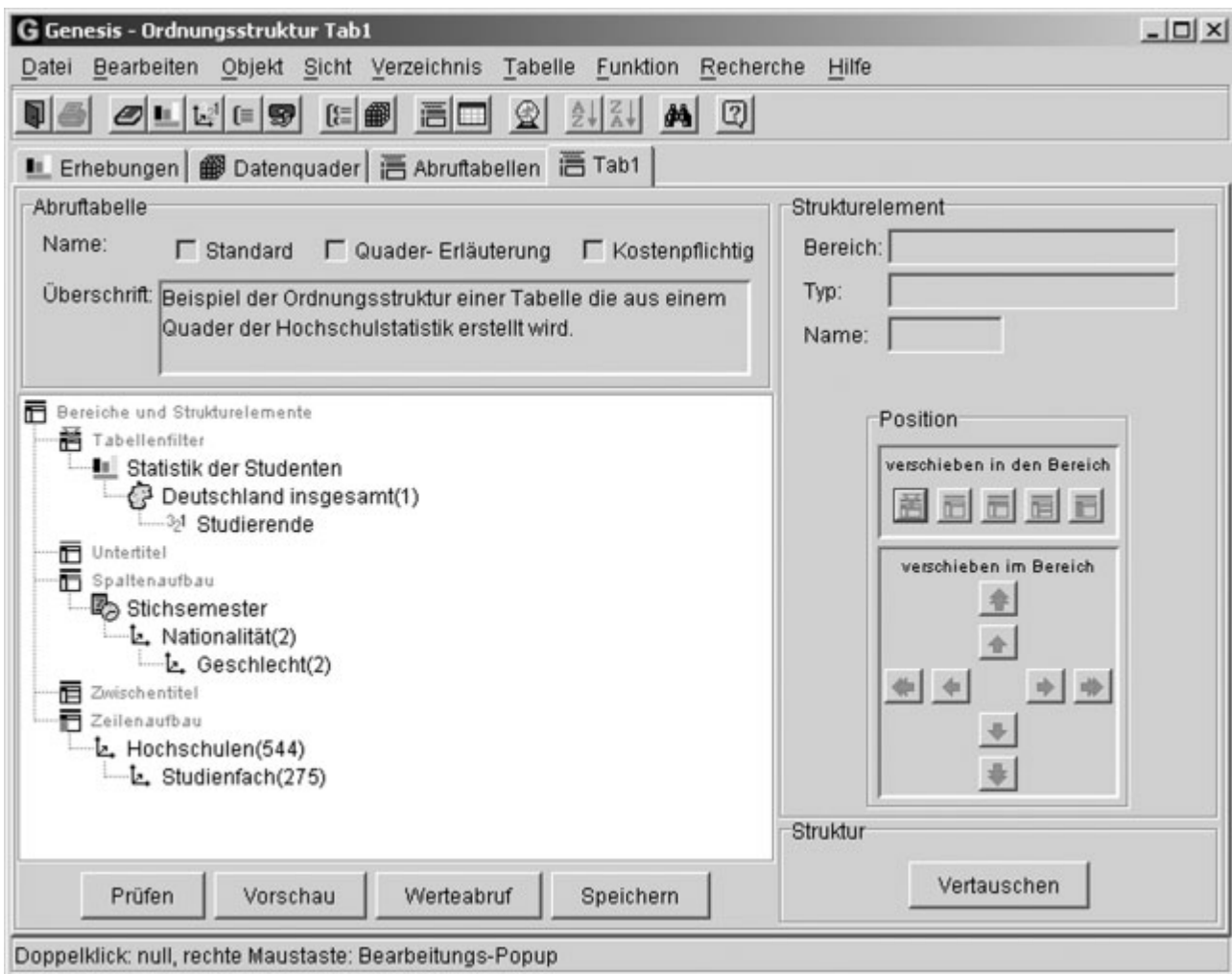


Tabelle 2: Beispiel für zwei verschiedene Tabellenrahmen mit unterschiedlichen Auswertungen aus dem gleichen Zahlenmaterial

Statistik der Studenten Deutschland Studierende (Anzahl)						
Hochschulauswahl Studienfachauswahl	Insgesamt	Geschlecht		Nationalität		
		männlich	weiblich	Deutsche	Ausländer	
WS 2003/04						
Uni Frankfurt am Main	Rechts- wissenschaft					
Statistik der Studenten Deutschland Studierende (Anzahl)						
Hochschulauswahl Studienfachauswahl	Nationalität					
	Insgesamt		Deutsche		Ausländer	
	Geschlecht			Geschlecht		
	männlich	weiblich	männlich	weiblich	männlich	weiblich
WS 2003/04						
Uni Frankfurt am Main	Rechts- wissenschaft					

Mit dem Tabellenassistenten steht jedem GENESIS-Nutzer ein flexibles Werkzeug zur Verfügung. Neben statistikbezogenen Auswertungen können durch die eindeutige Metadatenstruktur statistikübergreifende Tabellierungen und Analysen vorgenommen werden. Die früher übliche Vortabellierung, die trotz umfangreicher Sammlungen nicht immer alle Fragestellungen erfüllen konnte, ist für die im Auskunftssystem gespeicherten Daten unnötig geworden.

5.2 Verknüpfung von Daten

Beim Tabellenabruf können Daten aus mehreren Quadern und aus unterschiedlichen Statistiken zusammengeführt werden, wenn sie einheitliche Klassifikationen für ihre Sachgliederungen verwenden; dies ist ein sehr wichtiger Aspekt, der bei der Konzeption der Inhalte immer berücksichtigt werden muss. Übergreifende Auswertungen und Querschnittsveröffentlichungen werden damit wesentlich erleichtert. Denkbar wäre beispielsweise eine vergleichende Betrachtung der Konjunktorentwicklung im Produzierenden Gewerbe und im Baugewerbe, eine Gegenüberstellung der Insolvenzverfahren und der Anzahl der umsatzsteuerpflichtigen Unternehmen nach Bundesländern oder – wie in Schaubild 5 angedeutet – eine Auswertung zu Kapazität und Auslastung von Justizvollzugsanstalten. Bei der Tabellierung werden hier in einem Arbeitsschritt die Daten der Bevölkerungsstatistik mit denen der Strafvollzugs- und Maßregelvollzugsstatistik zusammengeführt.

Auswertungen der Daten nach Bundesländern können mit Hilfe von Regeln innerhalb von GENESIS durchgeführt werden; daneben ist die Berechnung auch in Tabellenkalkulationsprogrammen nach dem Export der Basistabellen möglich. Die Arbeitserleichterung durch die schnelle Bereitstellung von Informationen aus unterschiedlichen Statistiken in einem Datenbanksystem ist offensichtlich.

6 Kompatibilität und Anwendungsgebiete

Innerhalb des GENESIS-Verbundes ist der Datenaustausch standardisiert und weitgehend automatisiert; darüber hinaus gibt es viele Schnittstellen zu anderen Systemen. Die Software-Basis der Forschungsdatenzentren des Bundes und Länder nutzt das anpassungsfähige Speicher- und Verzeichnissystem von GENESIS. Gegenseitige Verweise und ein Austausch der Metadaten finden damit innerhalb eines Formats statt. Ein geplanter zentraler Nomenklaturserver⁶⁾ wird voraussichtlich ebenfalls eine Verbindung zu GENESIS haben und die Texte der standardisierten Klassifikationen an das Informationssystem übergeben.

6.1 Aufbereitung des Materials

Die Bundesstatistiken werden in der Regel von den Statistischen Landesämtern erhoben, nur bei ausgewählten Sta-

tistiken (Außenhandel) werden die Daten von den Auskunftgebenden direkt an das Statistische Bundesamt geliefert. Die Statistischen Ämter des Bundes und der Länder haben gemeinsame Datenbankinhalte abgestimmt. Der im statistischen Verbund geregelte Datenaustausch wird sukzessive um Importdateien für GENESIS erweitert, die damit Teil des statistischen Aufbereitungsprozesses werden.

Handelt es sich beim eingehenden Material um dezentral von den Landesämtern aufbereitete Daten, so werden die vereinbarten Importdateien für GENESIS per Datenfernübertragung an das Statistische Bundesamt geliefert. Bei den zentralen Statistiken werden die Statistischen Landesämter mit GENESIS-Importdateien beliefert.

Die statistische Geheimhaltung führt durch vorgeschaltete Programme zu einer Kennzeichnung der einzelnen Werte. Die Einspeicherung von primär oder sekundär geheim zu haltenden Daten und gegebenenfalls von Ersatzwerten ist vorgesehen. Abhängig von den Zugriffsrechten stehen die Werte zur weiteren Verarbeitung zur Verfügung oder werden bei nicht autorisierten Nutzern für die Anzeige gesperrt. Die Prüfung erfolgt über alle Hierarchiestufen hinweg. Als standardisiertes Programm ist eine Variante des vom Statistischen Bundesamt mitentwickelten ARGUS-Verfahrens vorgesehen, das so genannte τ -ARGUS für Tabellengeheimhaltung.⁷⁾ Darin wurden ein Verfahren zur iterativen Durchführung der Sekundärsperre (HITAS) sowie das Quaderverfahren des Landesamtes für Datenverarbeitung und Statistik Nordrhein-Westfalen integriert.

6.2 Export und Weiterverarbeitung

Der Zugriff auf die Daten in GENESIS erfolgt entweder direkt über die Quader oder in Form von Tabellen. Abhängig von den individuellen Zugriffsrechten können die Daten in ihrem Speicherformat aus dem System exportiert und diese Quader extern weiterverarbeitet werden. Dabei ist es möglich, die Erläuterungstexte und Definitionen mit herauszuziehen oder sich auf die notwendigen Metadaten zur Zuordnung der Werte zu beschränken. Dieser Export ganzer Quader ist für tiefgreifende Analysen der Datenbestände, zum Beispiel mittels SAS, möglich.

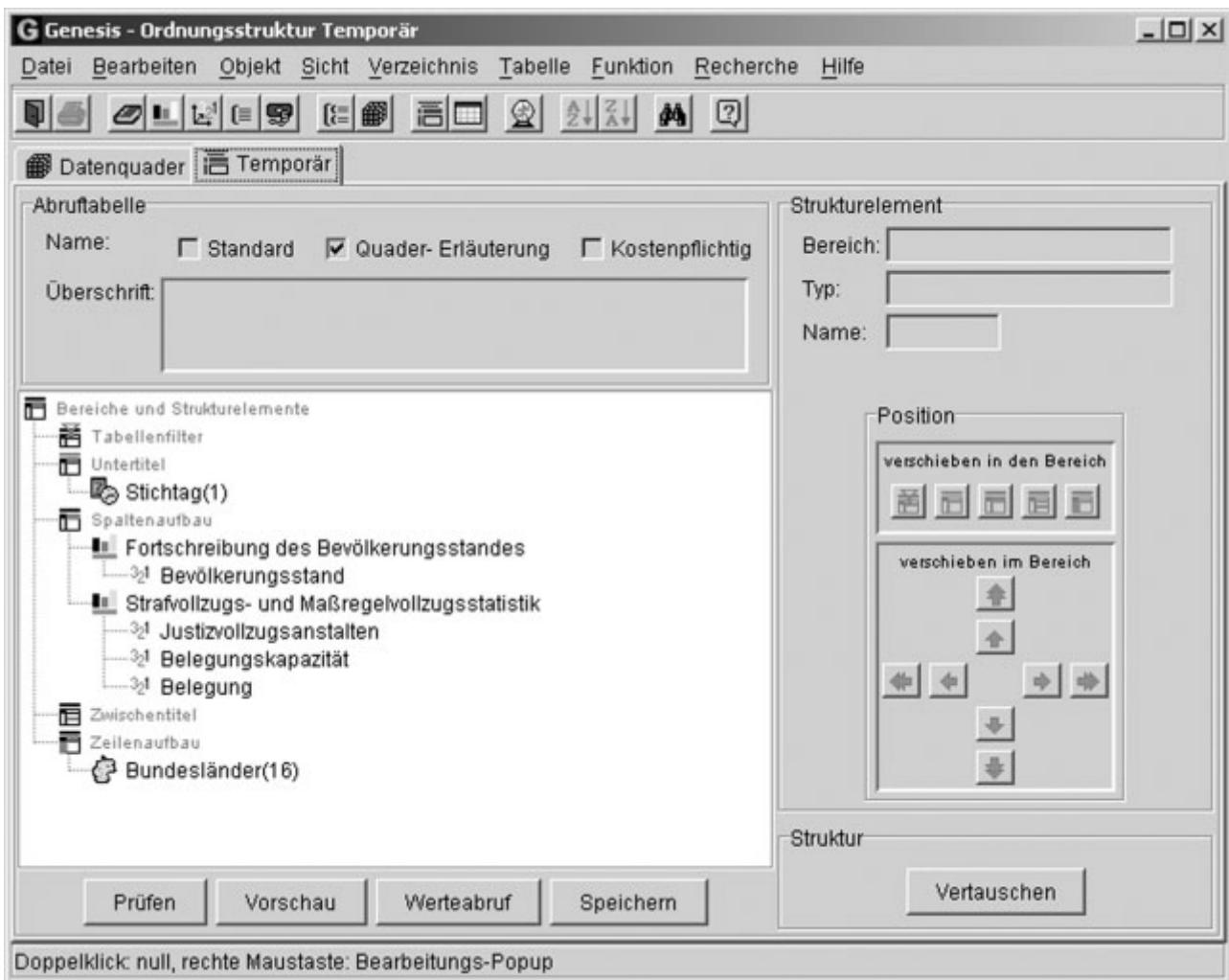
In der Regel wird GENESIS als Auskunftssystem genutzt, das heißt die gewünschten Daten werden in Form übersichtlicher Tabellen abgerufen. Das Produkt Tabelle tritt dabei in zwei Formen auf: als Endprodukt für die interne oder externe Präsentation statistischer Ergebnisse und als vorwiegend internes oder technisches Zwischenprodukt.

Für interne Zwecke gibt es die Möglichkeit, Daten im TabML-Format zu exportieren. TabML (Table Markup Language) ist eine auf XML 1.0 (Extensible Markup Language) basierende Auszeichnungssprache für Tabellen. Es handelt sich um eine Eigenentwicklung des Statistischen Bundesamtes primär für den Einsatz in der amtlichen Statistik. TabML ist aber so allgemein und flexibel formuliert, dass sie grund-

⁶⁾ Eine erste vom Bayerischen Landesamt für Statistik und Datenverarbeitung realisierte Version steht im gemeinsamen Statistik-Portal im Internet (<http://www.statistik-portal.de>) zur Verfügung.

⁷⁾ Siehe Giessing, S.: „Methoden zur Tabellengeheimhaltung in τ -ARGUS“ in WiSta-Sonderausgabe ISI-Weltkongress 2003, S. 66 f.

Schaubild 5



Doppelklick: null, rechte Maustaste: Bearbeitungs-Popup

Bundestländer	Fortschreibung des Bevölkerungsstandes	Strafvollzugs- und Maßregelvollzugsstatistik			Zusatzauswertungen:	
	Bevölkerungsstand	Justizvollzugsanstalten	Belegungs-kapazität	Belegung	Belegungs-quote	Kapazität zur Bevölkerung
	Anzahl	Anzahl	Anzahl	Anzahl	Prozent	Plätze pro 10 000 Einwohner
31.12.2002						
Schleswig-Holstein	2816507	6	1580	1401	88,7	5,6
Hamburg	1728806	11	3128	2673	85,5	18,1
Niedersachsen	7980472	19	6560	5907	90,0	8,2
Bremen	662098	1	826	587	71,1	12,5
Nordrhein-Westfalen	18076355	37	18360	14787	80,5	10,2
Hessen	6091618	17	5676	5505	97,0	9,3
Rheinland-Pfalz	4057727	10	3835	3440	89,7	9,5
Baden-Württemberg	10661320	20	8029	7729	96,3	7,5
Bayern	12387351	36	11522	11182	97,0	9,3
Saarland	1064988	3	848	776	91,5	8,0
Berlin	3392425	10	5022	4944	98,4	14,8
Brandenburg	2582379	7	2534	2165	85,4	9,8
Mecklenburg-Vorpommern	1744624	6	1624	1443	88,9	9,3
Sachsen	4349059	10	4382	4011	91,5	10,1
Sachsen-Anhalt	2548911	9	2787	2573	92,3	10,9
Thüringen	2392040	7	1753	1854	105,8	7,3

sätzlich überall dort Verwendung finden kann, wo Tabellen maschinell erzeugt oder gelesen werden.⁸⁾ Darin werden alle Informationen über die Struktur und den Inhalt (Metadaten) mitgegeben, um sie unabhängig von anwendungsspezifischen Formaten zur Weiterverarbeitung zur Verfügung zu stellen.

Für den internen Austausch von TabML gibt es Schnittstellen zu den beiden anderen Standardprodukten im Statistischen Bundesamt, SAS und StatSpez, und auch einen Konverter zur nachträglichen Ausgabe in die üblichen PC-Formate, die TabML-Workbench. Damit sind auch Daten aus GENESIS universell nutzbar.

7 Ausblick

Das statistische Informationssystem GENESIS ist im Statistischen Bundesamt erfolgreich eingeführt worden. Neben der öffentlichen Verbreitung des für externe Nutzer zugänglichen Teils über GENESIS-Online und der internen Nutzung durch die Informationsdienste im Hause haben sich noch zahlreiche weitere Anwendungsgebiete erschlossen. Sekundärstatistiken nutzen den zentralen Datenpool als Quelle. Laufende Querschnittsveröffentlichungen wie der Tabellenanhang der Zeitschrift *Wirtschaft und Statistik*, die Statistischen Wochenberichte oder die Tabellenübersichten der Website werden mit Hilfe von Prozeduren aus der Datenbank generiert. Jeder Autor von Publikationen im Statistischen Bundesamt kann sich der zentralen Datenbank bedienen, um Querschnittsbetrachtungen anzustellen.

Auch andere Behörden und Institutionen wie der Sachverständigenrat zur Begutachtung der gesamtwirtschaftlichen Entwicklung können über Sondernutzungsrechte über das behördeninterne TESTA-Netz den direkten Zugang zur Produktionsdatenbank erhalten. Mit der Deutschen Bundesbank funktioniert der Datenaustausch inzwischen in beide Richtungen – die Bundesbank ruft Originalwerte ab und liefert saisonbereinigte Daten für GENESIS.

Im statistischen Verbund nimmt das Interesse an GENESIS im Zuge der Möglichkeiten, die der Client und GENESIS-Online bieten, zu. So ist geplant, die jährlich erscheinende Regional-Datenbank auf CD-ROM, „Statistik regional“, periodisch aus GENESIS zu aktualisieren. Wenn das System selbst (noch) nicht in allen Statistischen Landesämtern eingesetzt wird, sind doch immer Schnittstellen zwischen den Datenbanken vorgesehen⁹⁾, das heißt Daten können nicht nur geliefert, sondern auch empfangen und genutzt werden.

Auch nach dem Projektabschluss wird in den beteiligten Statistischen Landesämtern und im Statistischen Bundesamt das System fachlich weiter ausgebaut und programmtechnisch gepflegt, um die wachsenden inhaltlichen und ergonomischen Anforderungen der Nutzer zu erfüllen. Es ist geplant, zusätzliche statistikspezifische Funktionen in

GENESIS zu integrieren, wie zum Beispiel die Berechnung von Veränderungsraten gegenüber Vorperioden.

Obwohl GENESIS auch als internes Auswertungswerkzeug zur Verfügung steht, ist die Hauptanwendung die Bereitstellung von Daten für externe Nutzer über GENESIS-Online. Diese über das Internet zugängliche Datenbank wird zur zentralen Auskunftsdatenbank des Statistischen Bundesamtes ausgebaut und soll das gesamte Spektrum der amtlichen Statistik Deutschlands abbilden. [u](#)

8) Nähere Informationen finden sich im Internet unter <http://www.statspez.de>.

9) Siehe dazu Wirtz, H.: „Das neue Landesinformationssystem des Statistischen Landesamtes“ in *Statistische Monatshefte Rheinland-Pfalz*, 7/2004, S. 181 ff.

Auszug aus Wirtschaft und Statistik

© Statistisches Bundesamt, Wiesbaden 2004

Für nichtgewerbliche Zwecke sind Vervielfältigung und unentgeltliche Verbreitung, auch auszugsweise, mit Quellenangabe gestattet. Die Verbreitung, auch auszugsweise, über elektronische Systeme/Datenträger bedarf der vorherigen Zustimmung. Alle übrigen Rechte bleiben vorbehalten.

Herausgeber: Statistisches Bundesamt, Wiesbaden

Schriftleitung: Johann Hahlen
Präsident des Statistischen Bundesamtes
Verantwortlich für den Inhalt:
Brigitte Reimann,
65180 Wiesbaden

- Telefon: +49 (0) 6 11/75 20 86
- E-Mail: wirtschaft-und-statistik@destatis.de

Vertriebspartner: SFG Servicecenter Fachverlage
Part of the Elsevier Group
Postfach 43 43
72774 Reutlingen
Telefon: +49 (0) 70 71/93 53 50
Telefax: +49 (0) 70 71/93 53 35
E-Mail: destatis@s-f-g.com

Erscheinungsfolge: monatlich



Allgemeine Informationen über das Statistische Bundesamt und sein Datenangebot erhalten Sie:

- im Internet: www.destatis.de

oder bei unserem Informationsservice
65180 Wiesbaden

- Telefon: +49 (0) 6 11/75 24 05
- Telefax: +49 (0) 6 11/75 33 30
- E-Mail: info@destatis.de