

Drei von vier Kunden der Deutschen Bahn waren 30 oder mehr Minuten unterwegs. Mit dem eigenen Pkw fuhren die meisten Pendler rund 10 bis 30 Minuten zur Arbeit. Aber auch 38% der Fahrradfahrer waren 10 bis 30 Minuten unterwegs. 10% der Erwerbstätigen gingen zu Fuß zur Arbeit, wovon 83% dann weniger als 10 Minuten benötigten. Weitere Informationen zu den genutzten Verkehrsmitteln und der damit benötigten Fahrtzeit sind Tabelle 2 zu entnehmen.

### Am Land hat's der ÖPNV schwer

Gerade in ländlichen Gegenden ist das Angebot des öffentlichen Personennahverkehrs nur selten ausreichend, um damit verhältnismäßig schnell zur Arbeit zu kommen. In Gemeinden unter 5000 Einwohnern nutzte nur jeder 17. Erwerbstätige Bus oder Bahn. In Städten mit 50000 bis unter 100000 Einwohnern war es schon jeder Zehnte, im Landesdurchschnitt jeder Neunte, der mit öffentlichen Verkehrsmitteln zur Arbeit fuhr. In den Großstädten Erlangen, Fürth, Ingolstadt, Regensburg und Würzburg entschied sich durchschnittlich jeder sechste erwerbstätige Einwohner für öffentliche Verkehrsmittel, in Augsburg und Nürnberg jeder Vierte, in München sogar jeder Dritte – wie aus Tabelle 3 hervorgeht.

### Der MVV in München wird angenommen

Die Daten für München, der einzigen Stadt Bayerns mit mehr als 500000 Einwohnern, weisen zum Teil erhebliche Unterschiede gegenüber den bayernweiten Durchschnittsdaten auf. Dabei wird aber nur das Pendlerverhalten der Münchner Bürger beschrieben; die Vielzahl derer, die von umliegenden Gemeinden des Münchner Verkehrs- und Tarifverbundes (MVV) in die Landeshauptstadt einpendeln, muß bei dieser Betrachtung wegen des Stichprobencharakters der Erhebung außer acht bleiben.

Nur 16% der Münchner waren in weniger als 10 Minuten an ihrem Arbeitsplatz, 48% benötigten zwischen 10 und 30 Minuten und 27% waren mehr als eine halbe Stunde unterwegs. Der Pkw war bei lediglich 37% der erwerbstätigen Münchner die erste Wahl als Verkehrsmittel für den Weg zwischen Wohnort und Arbeitsplatz. Annähernd genauso viele Erwerbstätige (35%) nutzten vorwiegend den MVV. Hinzu kommt ein unbekannter Anteil von Nutzern der Eisenbahn im Nahverkehr (MVV-Bereich). Weitere 7% benutzten das Fahrrad und 8% gingen zu Fuß. Diese Daten und die gesamt-bayerischen Vergleichswerte sind Tabelle 3 zu entnehmen.

Dipl.-Stat. Volker Birmann

## GENESIS ante portas

*Seit dem ersten Beitrag über GENESIS – das **G**emeinsame **N**eu(e) **S**tatistische **I**nformationssystem – im November 1995 hat das Projekt große Fortschritte gemacht. Die Entwicklung der Software ist weit vorangeschritten. Ein Großteil der Ziele ist erreicht. Dieser Beitrag berichtet über den Stand der Arbeiten, gibt einen kurzen Abriss über die Chronologie der Ereignisse und einen Überblick über die bei GENESIS beteiligten Gremien. Er geht ferner auf technische Aspekte des Systems ein und stellt einige der realisierten Funktionalitäten kurz vor. Auch der inhaltlichen Seite ist ein Abschnitt gewidmet.*

### Bisherige Entwicklung von GENESIS

Der Entschluß, ein gemeinsames neues statistisches Informationssystem für alle Statistischen Ämter im Verbund zu entwickeln, fiel Mitte 1991. Das Grobkonzept wurde von einer Bund/Länderarbeitsgruppe bis Mai 1992 entwickelt.

Für die weitere Entwicklung des Projekts wurde das „Kernteam GENESIS“ unter Beteiligung der Statistischen Landesämter Baden-Württemberg, Bayern, Berlin, Brandenburg, Niedersachsen und Sachsen sowie des Statistischen Bundesamts gebildet.

Für das Projektmanagement ist das Statistische Bundesamt zuständig. Die Projektleitung liegt beim Ausschuß „Organisationsfragen“ – einem Gremium auf Amtsebene. Auch die Amtsleiter selbst lassen sich regelmäßig über den Fortgang des Projektes berichten und entscheiden in wichtigen Punkten über das weitere Vorgehen.

In Zusammenarbeit mit einem vom Statistischen Landesamt Baden-Württemberg beauftragten externen Partner wurde im ersten Halbjahr 1993 das fachliche Feinkonzept entwickelt.

Ab Ende 1993 erarbeitete das „Kernteam GENESIS“ zusammen mit einem weiteren externen Auftragnehmer das DV-Konzept. Dieser programmierte auch die sog.

„Basisversion“, die 1995 übergeben wurde. Eine ausführliche Schilderung dieser Projektabschnitte findet sich in [1].

Vier Entwicklungsteams in Baden-Württemberg, Bayern, Brandenburg und Sachsen entwickeln seither GENESIS weiter. Die Länder Niedersachsen und Berlin haben sich mit Beginn der Programmierarbeiten in den Entwicklungsteams zurückgezogen. Das „Entwicklungsbüro“ im Statistischen Bundesamt koordiniert die umfangreichen Arbeiten und hat zudem die schwierige Aufgabe, die dezentral von den einzelnen Teams entwickelten Programmteile zu einer neuen GENESIS-Version zusammenzuführen.

Um die Fertigstellung der einzelnen Entwicklungsabschnitte zu beschleunigen, wurden einzelne Funktionalitäten auch extern vergeben. Das Landesamt Bayern gab den Export von Datenquadern und Metadaten (Erhebungen, Merkmale und Ausprägungen, Erhebungsmerkmale) in Auftrag, das Statistische Bundesamt die Funktionalität „Regeln“ – also die Ableitung von Merkmalen aus anderen Merkmalen.

Eine erste, funktional noch eingeschränkte Version „GENESIS 1.0“ wurde im Juli 1997 an interessierte Statistische Landesämter ausgeliefert. Viele wichtige und nützliche Funktionalitäten sind seitdem hinzugekommen, Menüs, Masken und Meldungstexte wurden überarbei-

tet, die Performance verbessert. Mit der Veröffentlichung dieses Beitrags ist GENESIS bei Version 1.6 angelangt.

### **Schwierigkeiten**

Die ursprünglichen Zeitpläne mußten – wie bei Projekten dieser Größenordnung nicht unüblich – im Laufe der Zeit mehrfach angepaßt werden. Verschiedene Einflußfaktoren ließen das Projekt langsamer vorankommen als geplant.

Die dezentrale Entwicklung erwies sich erwartungsgemäß als schwierig. Die Zuständigkeiten für die Funktionalitäten in GENESIS – z. B. Datenimport/-export, Datenschichtoperationen, Thesaurus, Tabellenabruf, Zugangs- und Zugriffsschutz – wurden auf die Entwicklungsteams verteilt. Trotz hoher Modularisierung gab und gibt es zwangsläufig Überschneidungen bei der Bearbeitung von Modulen. In Anbetracht der räumlichen Trennung (München, Kamenz, Potsdam, Stuttgart, Wiesbaden) ist die Abstimmung zwischen den Programmerteams nicht leicht. Auch die Zusammenführung des neuen, in den vier Teams für die nächste Version erstellten Programmcodes – die sog. „Integration“ – ist durch die genannten Überschneidungen eine schwierige Aufgabe. Sind bei einem Team Erweiterungen oder Änderungen von Datenbankdateien, Schnittstellen und Masken notwendig, müssen alle anderen Teams „ihre“ Module ggf. entsprechend anpassen. Und schließlich muß auch noch getestet werden, ob nach dem „Einklinken“ neuer Funktionalitäten in bestehende Programmteile auch die „alten“ noch fehlerfrei funktionieren.

Gerade in einem Projekt dieser Größenordnung erweisen sich auch Personalfuktuationen als „Ressourcenfresser“. Der Einarbeitungsaufwand für den „Quereinsteiger“ ist enorm und zieht zudem Kapazitäten bei den eingearbeiteten Kollegen ab. Umfangreiches Wissen ist notwendig, um in diesem Projekt produktiv mitarbeiten zu können. Schließlich müssen zahlreiche technische Konventionen eingehalten werden, um ein reibungsloses Zusammenwirken sowie ein einheitliches Erscheinungsbild und Systemverhalten zu gewährleisten. Dies setzt vertiefte Kenntnisse der Konzepte voraus: Datenmodell, Programmierkonventionen, Funktionsweise der Entwicklungsumgebung, Zusammenwirken der Programmkomponenten, Dialoggestaltung und -abläufe (festgelegt in einem Style-Guide), Konventionen der Dokumentation, organisatorische Abläufe usw.

Auch das zeitweise Abziehen von Entwicklern für andere Aufgaben verringerte die Entwicklungskapazitäten und verhinderte im Verlauf des Projekts die Fertigstellung einzelner Aufgaben zu dem im Projektplan vorgesehenen Termin.

### **Beteiligung der Anwender**

Um GENESIS auch aus der Anwenderperspektive zu beleuchten, setzten die Amtsleiter im Mai 1997 die „GENESIS-Nutzergruppe“ ein. Sie besteht derzeit aus Nutzern der Länder Baden-Württemberg, Brandenburg, Niedersachsen, Nordrhein-Westfalen, Rheinland-Pfalz, Sachsen und Thüringen. Die Federführung liegt beim Statistischen Bundesamt.

Den ersten Bericht über den Stand der Arbeiten, gewonnene Erfahrungen und Verbesserungsvorschläge hat die Gruppe Anfang 1998 vorgelegt und danach kontinuierlich ergänzt. Die erarbeiteten Änderungswünsche und Verbesserungsvorschläge wurden im Kernteam auf ihre

Machbarkeit hin geprüft, die Realisierungsaufwände ermittelt und die neue Aufgabe entsprechend ihrer vereinbarten Priorität in die Arbeits- und Projektplanung integriert.

In mehreren gemeinsamen Workshops des Kernteam und der Nutzergruppe wurden Änderungswünsche, Verbesserungsvorschläge und Konzepte diskutiert, Prioritäten vergeben und Erfahrungen ausgetauscht. Für die Tests werden den beteiligten Ländern die jeweils neuesten integrierten GENESIS-Versionen zur Verfügung gestellt. Kontinuierlich werden so die neuen Versionen getestet und die GENESIS-Entwicklung aus Anwendersicht begleitet.

### **Systemarchitektur**

GENESIS basiert auf dem Datenbanksystem ADABAS und der Programmiersprache NATURAL der Software AG. Diese Software ist in allen Statistischen Ämtern der Länder und des Bundes verfügbar. Damit ist GENESIS auf den Betriebssystemen OS/390 (IBM), OSD (Siemens) und Unix einsetzbar.

GENESIS ist stark modularisiert, um die dezentrale Entwicklung zu ermöglichen und eine Client/Server-Version mit grafischer Benutzeroberfläche konzeptionell zu unterstützen. Grundfunktionen, Anwendungsfunktionen sowie Funktionen für den Datenzugriff und die Dialogsteuerung sind jeweils in eigenen Schichten zusammengefaßt.

In der vorliegenden Ausbaustufe bietet GENESIS eine zeichenorientierte Oberfläche für die Nutzung über Terminals bzw. PC mit Host-Emulation. Eine Client/Server-Version mit grafischer Oberfläche wird derzeit vom Statistischen Bundesamt entwickelt. Der Client – sozusagen die ausgegliederte Benutzeroberfläche auf dem PC – greift dabei auf den GENESIS-Server auf einem Großrechner bzw. einem Unix-System zu. Die plattformübergreifende Kommunikation der Anwendungssoftware ermöglicht EntireX, eine sog. Middleware der Software AG. Der Client selbst ist in Java programmiert.

In der zeichenorientierten Oberfläche des Großrechners wurden so weit möglich, die Dialogelemente einer Windowsoberfläche nachempfunden. So gibt es beispielsweise eine cursorsensitive Menüleiste: positioniert man den Cursor auf einen Menüeintrag und drückt ENTER, klappt das Menü herunter und zeigt alle in diesem Menüpunkt möglichen Funktionen. Auch wurden in Anlehnung an grafische Oberflächen Dialogboxen für Meldungen, Rückfragen und die Eingabe von Parametern realisiert. Wahlmöglichkeiten werden auf den Masken ähnlich den in Windows üblichen Checkboxes und Optionsfeldern angeboten. Für die Auswahl von Objekten – Erhebungen, Merkmale, Merkmalsausprägungen, Regeln etc. – wird über das Menü „Sicht“ in das entsprechende Verzeichnis verzweigt. Dort können die Objekte durch „Ankreuzen“ ausgewählt werden. Der Zeilenbefehl „?“ vor einem Objekt öffnet ein PopUp-Fenster mit den aktuell verfügbaren Funktionen. Er übernimmt damit die Funktion der rechten Maustaste in einem Windowsystem.

Ein sog. „Kellerungsmechanismus“ simuliert die Fenster-technik. So ist es z. B. möglich, Informationen in verschiedenen Verzeichnissen nachzuschlagen, während man eine Abruftabelle erstellt und schrittweise zum vorherigen „Fenster“ zurückzukehren. Durchaus keine Selbstverständlichkeit in der Großrechnerwelt.

Außerdem gibt es eine Hilfefunktion. Mit der „F1“-Taste wird ein Fenster mit Hilfetexten zur aktuellen Maske geöffnet. Die „F2“-Taste informiert über die Belegung der Funktionstasten.

### Das Datenmodell im Überblick

Wertedaten werden in GENESIS in Form mehrdimensionaler Datenkörper – den sog. Datenquadern – abgebildet und durch Metadaten beschrieben (siehe auch [1]). Sie enthalten Werte jeweils einer Erhebung. Klassifizierende Merkmale mit ihren Ausprägungen (z.B. das Merkmal „Geschlecht“ mit seinen Ausprägungen „männlich“, „weiblich“) und das identifizierende Zeitmerkmal (z.B. „Stichtag“) beschreiben die Achsen und die Achsenkoordinaten des Datenquaders. Die Wertmerkmale, wie z.B. „Bevölkerung“, „Umsatz“, „Wohngebäude“, definieren zusammen mit der Maßeinheit (z.B. „Anzahl“, „ha“, „Tonnen“) die Datenquaderinhalte. So ist ein Wert an den Schnittpunkten der Koordinaten (z.B. „Gestorbene“, „weiblich“, „Niederbayern“, „2000“) sachlich, räumlich und zeitlich eindeutig definiert (siehe Abbildung 1).

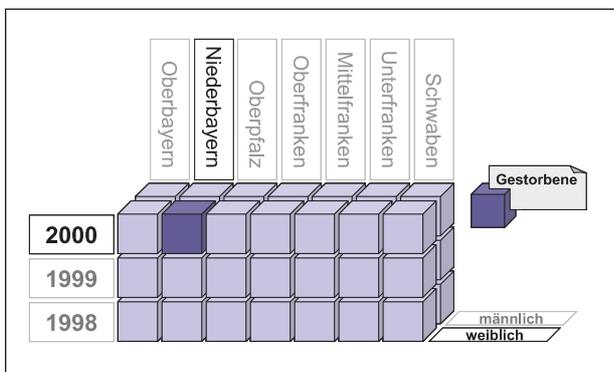


Abbildung 1

Die Metadaten werden in Verzeichnissen zur Recherche angeboten. Die für den Benutzer wichtigsten Verzeichnisse enthalten die Informationen über

- Begriffe (Schlagworte)
- Erhebungen
- Merkmale
- Merkmalsausprägungen
- Datenquader
- Regeln
- Maßeinheiten
- Abruf- und Standardtabellen
- Ergebnistabellen.

Aus den Texten der beschreibenden Metadaten werden nach der Eingabe bzw. Korrektur sowie beim Import automatisch Begriffe extrahiert und im Thesaurus verwaltet. Die Begriffe sind mit dem Objekt verknüpft, aus dem sie gebildet wurden. Damit ist eine Recherche auch über Schlagworte möglich.

Darüber hinaus sind natürlich auch Systemdaten wie z.B. Benutzer und Benutzergruppen sowie deren Zugriffsberechtigungen in Form der sog. Zugriffsprofile vorhanden.

### Recherchieren in GENESIS

Die Recherche soll dem Nutzer dabei helfen, möglichst einfach und schnell die Informationen zu finden, die er benötigt. GENESIS bietet hierzu zwei Möglichkeiten:

- der Einstieg über Verzeichnisse
- die Suche über Stichworte.

In den Verzeichnissen kann sich der Nutzer über die vorhandenen Datenobjekte – Erhebungen, Merkmale usw. – informieren. Das Menü „Ansicht“ bietet hierbei verschiedene Möglichkeiten, die angezeigten Daten zu sortieren. Über Filter kann die Datenmenge gezielt reduziert werden. So kann man sich beispielsweise im Merkmalsverzeichnis nur Regionalmerkmale anzeigen lassen oder auch alle in einer Begriffsrecherche ermittelten Merkmale. Auch die Anzeige aller Einträge, die einen bestimmten Text enthalten, ist über die Selektionsfunktion möglich. So findet beispielsweise die Eingabe „Alter“ im Merkmalsverzeichnis alle Merkmale, deren Kurztext mit „Alter“ beginnt, der Suchbegriff „\*Alter\*“ sucht den Text „Alter“ im gesamten Kurztext.

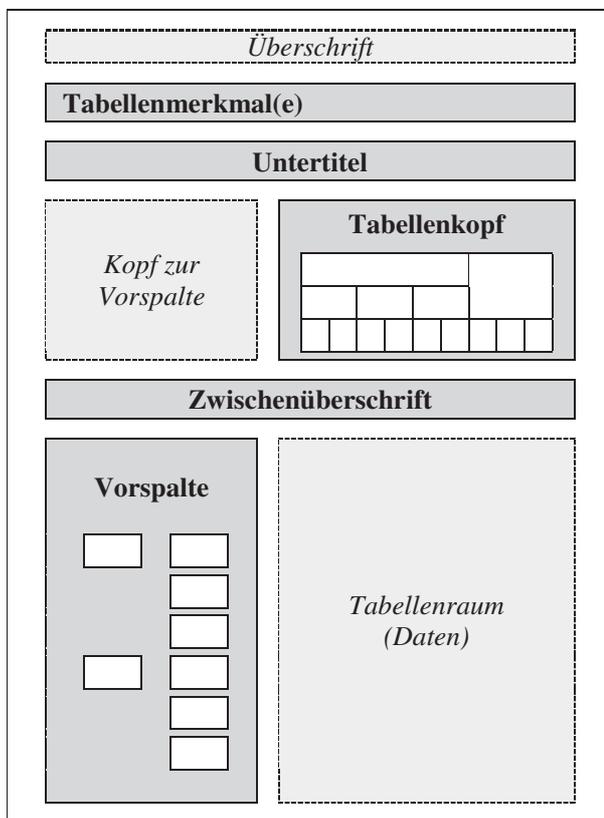


Abbildung 2

Zu den Objekten selbst kann man deren Eigenschaften abfragen. Bei Merkmalen beispielsweise den Merkmals-typ (räumliche, sachliche, zeitliche Klassifikation), den Gliederungstyp („davon“, „darunter“, „und zwar“), die Standardsortierung (Sortierung nach Fachschlüssel, Kurztext, Positionsnummer), aber auch Langtext, Notizen und Definition. Letztere gibt es auch für Merkmalsausprägungen und Maßeinheiten. Für Erhebungen sind Erläuterungen abrufbar.

Bestehen Beziehungen zu Objekten in anderen Verzeichnissen, so ist ein direktes Verzweigen möglich. Beispielsweise können die Merkmalsausprägungen eines

Merkmale abgefragt, die Erhebungsmerkmale einer Erhebung angelistet oder die zu einer bestimmten Erhebung verfügbaren Datenquader gesucht und angezeigt werden. Auch die Anzeige der verfügbaren Abruftabellen zu einer Erhebung ist vorgesehen. So ist ein Navigieren zwischen den Verzeichnissen möglich.

Über die im Thesaurus verwalteten Begriffe findet man durch Eingabe von Schlagworten ohne Kenntnis über die Struktur der Daten die gewünschten Informationen im System.

Bei einfachen Begriffsrecherchen ist es möglich, durch Eingabe weniger Suchbegriffe direkt einen Werteabruf zu starten und so zu den benötigten Zahlen zu gelangen. Dabei wird in Zwischenschritten die Ergebnismenge angezeigt und im Dialog sukzessive verfeinert oder erweitert. Passen mehrere Datenquader zu den eingegebenen Suchbegriffen, werden sie zur Auswahl angezeigt. Als nächstes werden die im gewählten Datenquader enthaltenen Wertmerkmale und schließlich die Achsen zur Auswahl angeboten, um die Menge der abzurufenden Werte zu erweitern oder einzuschränken. Entsprechend der eingegebenen Suchbegriffe werden vom System bereits Objekte vorselektiert. Nach einer Auswahl der gewünschten Jahre startet der Werteabruf und ermittelt die gewünschten Zahlen.

### Möglichkeiten des Werteabrufs

Hauptziel eines Statistischen Informationssystems ist es natürlich, Daten in verschiedenen Sichten, Aggregationen und Kombinationen zur Verfügung zu stellen. In GENESIS ermöglicht dies der Werteabruf. Inhalt und Struktur der gewünschten Tabelle werden in der sog. Abruftabelle definiert. Dabei werden den Strukturelementen einer Tabelle (z.B. Vorspalte, Tabellenkopf, siehe Abbildung 2 und Abbildung 3 bis 6) Merkmale zugeordnet. Ein besonders schneller und komfortabler Weg zur Abruftabelle führt über den Tabellenassistenten, der in einem Dialog aus einem Datenquader eine Abruftabelle erstellt. Dabei kann der Benutzer Schritt für Schritt wählen, welche Wertmerkmale, Achsen und Zeiträume des Datenquaders er benötigt (siehe Beispiel am Ende des Beitrags).

Jedes der Strukturelemente hat bestimmte Eigenschaften. Die Abbildungen 5 bis 8 zeigen eine Tabelle aus der Geburtenstatistik, in der die verschiedenen Strukturelemente verwendet werden.

Die in einem *Tabellenmerkmal* definierte Vorbedingung gilt für die gesamte Tabelle. Gibt man hier ein Wertmerkmal an, werden in der Tabelle nur Werte zu genau diesem Merkmal ausgewiesen. Bei klassifizierenden oder zeitidentifizierenden Merkmalen muß genau eine Ausprägung ausgewählt werden, so z.B. die Ausprägung „verheiratet“ des Merkmals „Familienstand“ oder das Jahr „2000“ des Merkmals „Berichtsjahr“. Mehrere Tabellenmerkmale werden mit einem logischen „und“ verknüpft und schränken die in der Tabelle nachgewiesenen Werte entsprechend ein. Die Ausgabe der Tabellenmerkmale erfolgt direkt unter der Tabellenüberschrift (siehe Wertmerkmal „Lebendgeborene“, Abbildung 6).

Der *Untertitel* ist als höchste Gliederungsebene einer Tabelle über dem Tabellenkopf angeordnet. Hier können beliebige klassifizierende oder zeitidentifizierende Merkmale angegeben werden. Für jede einzelne Ausprägung eines Merkmals im Untertitel wird bei der Ausgabe auf dem Terminal bzw. beim Ausdruck eine neue Seite be-

gonnen. In der Beispieltabelle wird für jeden Kreis eine neue Seite begonnen.

Im *Tabellenkopf* und in der *Vorspalte* können Merkmale sehr komplex angeordnet werden. Dabei sind zwei grundsätzliche Darstellungen möglich:

- Ein Merkmal wird durch ein weiteres Merkmal zusätzlich untergliedert. Das Merkmal wird dem ersten quasi „untergeordnet“. Die Tabelle weist somit alle Ausprägungskombinationen aller untergeordneten Merkmale nach:

männlich		weiblich	
Deutsche	Ausländer	Deutsche	Ausländer

Ein Merkmal, das einem anderen untergeordnet wird, ist in der Abruftabelle eingerückt dargestellt (siehe Merkmale „Geschlecht“ und „Nationalität“ in Abbildung 3 und Ergebnistabelle in Abbildung 4).

- Die Ausprägungen von Merkmalen werden gleichwertig „nebeneinander“ dargestellt, jedoch *nicht* kombiniert:

männlich	weiblich	Deutsche	Ausländer
----------	----------	----------	-----------

(siehe Abbildung 5 und 6)



Abbildung 3

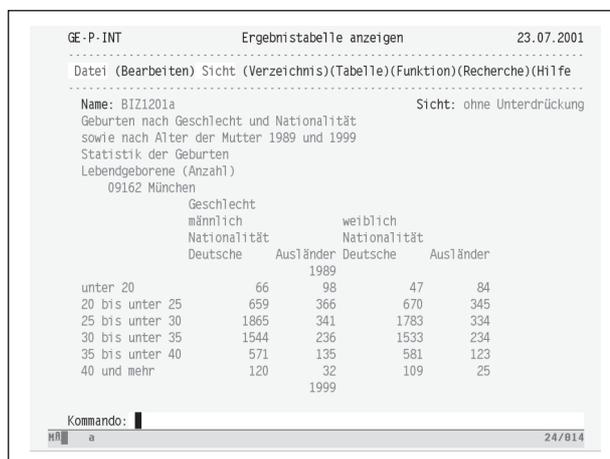


Abbildung 4



Abbildung 5

	Geschlecht		Nationalität	
	männlich	weiblich	Deutsche	Ausländer
1989				
unter 20	164	131	113	182
20 bis unter 25	1025	1015	1329	711
25 bis unter 30	2206	2117	3648	675
30 bis unter 35	1780	1767	3077	470
35 bis unter 40	706	704	1152	258
40 und mehr	152	134	229	57
1999				
unter 20	102	127	109	120
20 bis unter 25	733	744	686	791

Abbildung 6

Unterordnungen und Nebenordnungen können auch kombiniert werden. So ist es sehr einfach möglich, verschiedene Datenquader zu „verschneiden“ – selbst dann, wenn sie aus unterschiedlichen Erhebungen stammen.

Die *Zwischenüberschrift* ist quasi die höchste Ebene einer Vorspaltengliederung. Die Texte der Merkmalsausprägungen werden dabei eingerückt als Zwischenüberschrift in der Tabelle ausgegeben (siehe Berichtsjahre 1999 und 1989 in Abbildung 6).

Der GENESIS-Nutzer kann bei den einzelnen Strukturelementen eine Vielzahl von Eigenschaften festlegen.

Bei klassifizierenden Merkmalen:

- Die Sortierung der Merkmalsausprägungen erfolgt wahlweise auf- oder absteigend, nach Fachschlüssel oder alphabetisch nach dem Kurztext, aber auch nach einer hinterlegten „üblichen“ Reihenfolge (wie z. B. „ledig“, „verheiratet“, „verwitwet“, „geschieden“), der sog. „Positionsnummer“.
- Wahlweise kann bei den Merkmalsausprägungen der Kurztext, der Fachschlüssel oder auch beides ausgegeben werden.
- Aus der verfügbaren Menge der Merkmalsausprägungen kann eine Auswahl getroffen werden.

– Auf Wunsch wird eine Ingesamtsumme für das klassifizierende Merkmal oder auch eine Summe über die getroffene Auswahl ausgegeben. Die Positionierung der Summe erfolgt optional vor oder nach den Merkmalsausprägungen.

– Ferner kann bei einem klassifizierenden Merkmal auch eine Referenz- oder Hierarchieregel angegeben werden. So ist es mit der entsprechenden Referenzregel beispielsweise möglich, die Gemeinden nach Arbeitsamtsbezirken anzuordnen und wahlweise auch entsprechende Aggregate zu bilden. Mit der Hierarchieregel können z. B. Gemeinden, Kreise, Regierungsbezirke und das Land Bayern entsprechend ihrer hierarchischen Struktur dargestellt werden.

Bei Zeitachsen ist eine Auswahl des Zeitraums in Form einer Einzelauswahl oder der Angabe von halboffenen oder geschlossenen Intervallen (z. B. jedes zweite Jahr von 1990 bis 2000) möglich. Besonders nützlich ist auch die Option, sich auf das aktuellste Jahr zu beziehen. So ist es z. B. möglich, daß eine Tabelle, ausgehend vom aktuell verfügbaren Jahr standardmäßig die jeweils letzten 10 Jahre ausgibt.

Ist die Definition der Tabelle abgeschlossen, kann der Wertabruf gestartet werden. Aus der Abruftabelle wird so die Ergebnistabelle.

Bevor die Daten jedoch abgerufen werden, findet ein komplexer Analyseprozeß statt, in dem GENESIS die Abruftabelle in Teiltabellen zerlegt, die jeweils *einem* logischen Datenquader entsprechen. Zu jeder dieser Teiltabellen wird ein Datenquader gesucht. Ist ein genau passender Datenquader gefunden, wird dieser verwendet. Gibt es jedoch nur einen Datenquader, der mehr Achsen hat, als in der Tabelle benötigt werden, wird eine ad hoc-Aggregation durchgeführt. Findet GENESIS keinen gespeicherten Datenquader, wird geprüft, ob mit Hilfe der im System vorhandenen Referenz-, Hierarchie- oder Algorithmusregeln ein Datenquader erzeugt werden kann.

### Standardtabellen

Noch einfacher gestaltet sich ein Datenabruf, wenn man auf die sog. Standardtabellen zurückgreift. Das sind vordefinierte Tabellen, die nur noch wenige, vom Ersteller vorgegebene Wahlmöglichkeiten bieten. Dies kann eine Zeitauswahl sein oder einzelne Ausprägungen eines klassifizierenden Merkmals. Es ist aber auch möglich, verschiedene Merkmale in der Tabelle zur Auswahl anzubieten; z. B. „Bevölkerung nach Geschlecht und Nationalität oder Familienstand“.

Die Standardtabellen können zum einen über Verzeichnisse recherchiert werden. So ist es möglich, sich aus dem Erhebungs- bzw. Merkmalsverzeichnis alle Abruf- und Standardtabellen einer bestimmten Erhebung bzw. eines Merkmals anzeigen zu lassen. Auch der Weg über die Eingabe von Schlagworten im Rahmen einer Begriffsrecherche führt schnell zum Ziel.

### Ausgabe und Export von Ergebnistabellen

Die Ergebnistabelle kann auf dem Bildschirm betrachtet, gedruckt oder für die Weiterverarbeitung bzw. die Lieferung im Rahmen des Auskunftsdienstes auch in verschiedenen Formaten exportiert werden.

Der Bildschirm eines Terminals beschränkt die Menge des darstellbaren Textes auf 25 Zeilen mit jeweils 80 Zei-

chen. Nach Abzug der Zeilen für die Dialogsteuerung („Fenstertitel“, Menüleiste) bleiben noch 18 Zeilen für die Ausgabe der Ergebnistabelle. Je mehr Text für die Beschriftung der Tabelle in Kopf und Vorspalte ausgegeben wird, desto weniger steht für die Wertedaten zur Verfügung. Dies ist eine Gratwanderung zwischen dem Anspruch nach verständlich beschrifteten Tabellen einerseits und dem Wunsch, möglichst viele Wertedaten auf einer Bildschirmseite darzustellen.

Die Ausgabe von Daten in Dateiform trägt den geänderten Anforderungen der DV-Landschaft Rechnung. Der PC hat seinen Siegeszug angetreten. Die Weiterverarbeitung der statistischen Daten erfolgt kaum noch auf dem Großrechner, sondern größtenteils auf dem PC. Zudem erlangt das Internet mit seinen Formaten eine immer größere Bedeutung. Deshalb wurde auf die Ausgabe aufwendiger und dialogintensiver Datensätze mit festem Aufbau zugunsten der universellen Dateiformate CSV und HTML verzichtet.

Der Export von Ergebnistabellen bietet deshalb in der vorliegenden Version folgende Datenformate:

- CSV (Comma separated values). Hier werden die Datenfelder in jedem Satz durch Semikolon getrennt. Es gibt zwei Varianten:
  - eine tabelleorientierte, die im Rahmen der Möglichkeiten des Formats die Tabellenstruktur in die Ausgabedatei übernimmt,
  - eine datensatzorientierte Variante, die jeden Datensatz mit allen Identifikatoren der Strukturelemente „Vorspalte“, „Untertitel“ und „Zwischentitel“ ausgibt, um den Inhalt jedes einzelnen Datensatzes eindeutig identifizieren zu können.
- RegioStat-Format: für das gemeinsame Datenangebot des Bundes und der Länder „Regio-Stat“<sup>2</sup> wurde ein einheitliches Datenformat vereinbart. Es ist im Grunde ein spezielles CSV-Format, das Reihenfolge und Format für die Datenfelder vorgibt: Tabellenummer, Zeitangabe, Regionalschlüssel, laufende Nummer der Tabellenzeile und schließlich die Wertfelder.
- HTML – das Standardformat für das Internet. Die Datei kann direkt im Internetbrowser dargestellt werden, eignet sich aber auch für die Übergabe an Textverarbeitungen und Tabellenkalkulationen, die in der Lage sind, HTML zu interpretieren. Die Formatierung, wie z.B. verbundene Zellen in Spaltenüberschriften, bleibt dabei erhalten.

In der PC/Client-Version ist zudem eine direkte Übergabe der Ergebnistabelle an Microsoft Excel möglich.

### Die inhaltliche Seite von GENESIS

Nicht nur der softwaretechnische Teil des gemeinsamen neuen statistischen Informationssystems soll mit GENESIS auf eine gemeinsame Grundlage gestellt werden. Auch die Inhalte, die in den bestehenden statistischen Datenbanken des Bundes und der Länder bislang stark voneinander abweichen, sollen so weit wie möglich harmonisiert werden. Mit dieser Aufgabe befaßt sich seit Juni 1993 die „Unterarbeitsgruppe (UAG) Datenbankinhalte“. In der UAG waren anfänglich nur die Statistischen Ämter Baden-Württemberg, Bayern, Berlin, Nordrhein-Westfalen, Saarland, Sachsen und Schleswig-Holstein vertreten. Später kamen noch Brandenburg, Bremen, Hessen, Niedersachsen, Rheinland-Pfalz und Thüringen

hinzu. Die interne Projektleitung wurde dem Statistischen Bundesamt übertragen.

Die bundesweite Abstimmung und Vereinheitlichung der Datenbankinhalte soll die arbeitsteilige Erstellung und Pflege der Daten vereinfachen und die Informationsangebote der einzelnen Statistischen Ämter weitgehend vereinheitlichen. Davon profitieren nicht nur die Konsumenten der amtlichen Statistik, sondern auch die Statistikproduzenten, d.h. die Statistischen Ämter. Das gemeinschaftliche Auftreten mit einem einheitlichen Erscheinungsbild am Informationsmarkt führt zu einem Image- und Akzeptanzgewinn, der letztlich der gesamten Statistik zugute kommt.

Im Rahmen des GENESIS-Projektes definiert die UAG für ausgewählte Erhebungen (in erster Linie für häufig nachgefragte Statistiken, zu denen bisher in mindestens sechs Datenbanken der Statistischen Ämter des Bundes und der Länder Ergebnisse gespeichert sind) ein gemeinsames Datenangebot in Form von Quadern, die in Zukunft in allen Datenbanken für vielfältige Auswertungs- und Abrufzwecke verfügbar sein sollen. In Anbetracht der sehr unterschiedlichen Struktur der Länder – das Spektrum reicht schließlich von Stadtstaaten bis zu den Flächenländern mit z.T. mehr als 2000 Gemeinden – eine nicht einfache und zeitaufwendige Aufgabe. Zudem erschweren bei vielen Statistiken Geheimhaltungsaspekte die Festlegung der Datenbankinhalte.

Die Quader liegen entweder als für alle Statistischen Ämter des Bundes und der Länder verbindliche Veröffentlichungsquader oder als fakultative Arbeitsquader vor, wobei erstere jedermann zugänglich zur Verfügung gestellt werden, während letztere nur für hausinterne Benutzer mit eigener Zugangsberechtigung zugänglich sind, da hier auch noch Geheimhaltungsfälle und methodische Probleme enthalten sein können. Die Veröffentlichungsquader müssen mindestens die Datenbestände der Mindestveröffentlichungsprogramme<sup>1)</sup> der Länder sowie des Regio-Stat-Kataloges<sup>2)</sup> abdecken. Die Arbeitsquader geben i.d.R. das Tabellenprogramm der Verbundprogrammierung wieder, wobei dies nicht für Großzählungen wie z. B. die Volkszählung gilt. Zudem werden in den Arbeitsquadern Sonderwünsche einzelner Landesämter, die durch die Verbundprogrammierung nicht abgedeckt werden, aufgenommen. Insgesamt hat die UAG rund 80 Statistiken wie z.B. die Wanderungsstatistik, die Gewerbeanzeigenstatistik, die Statistik der Baugenehmigungen und die Statistik der Baufertigstellungen, die Krankenhausstatistik, die Wahlstatistiken usw. ausgewählt, die in allen GENESIS-Datenbanken der Statistischen Ämter abrufbar sein sollen. Vom Bayerischen Landesamt wurde die Definition einheitlicher GENESIS-Datenquader federführend für die Volkszählung 1987, die Arbeitsstättenzählung 1987, die Beschäftigtenstatistik, das Gemeindeverzeichnis und die Feststellung des Gebietsstandes übernommen. In 23 UAG-Sitzungen wurden bislang rund 60 Statistiken mit bundeseinheitlich abgestimmten Datenquadern verabschiedet, die derzeit spezifiziert und programmiert werden. Die Datenquader der Arbeitsstättenzählung 1987, der Volkswirtschaftlichen Gesamtrechnungen der Länder und der Flächenenerhebung stehen u.a. für GENESIS zur Verfügung.

Eine zweite inhaltsorientierte Gruppe – die „Unterarbeitsgruppe Metadaten“ – beschäftigt sich seit April 1995 mit der Thematik der Metadaten in GENESIS. Den Vorsitz in der Metadatengruppe hat das Statistische Bundesamt,

das eine koordinierende Funktion bei der Erstellung der Metadaten übernommen hat. Darüber hinaus setzt sich die Metadatengruppe noch aus Vertretern der Statistischen Ämter Baden-Württemberg, Niedersachsen, Rheinland-Pfalz und Sachsen zusammen. Bayern beteiligte sich zeitweise im Rahmen eines Pilotprojektes aus dem Bereich der Bevölkerungsstatistik an dieser UAG.

Bei den Metadaten handelt es sich um vollständige, aktuelle und standardisierte Beschreibungsinformationen, die einen Wert räumlich, sachlich und zeitlich eindeutig definieren. Im einzelnen geht es hierbei u.a. um Angaben zu Erhebungen, zu Merkmalen und ggf. deren Merkmalsausprägungen, zu Regeln, zu Maßeinheiten und zu Datenquadern. Während der Test- und Aufbauphase wurde beschlossen, daß die Metadaten zentral vom Statistischen Bundesamt verwaltet werden. Es soll die von den Statistischen Landesämtern vergebenen Metadaten sammeln und darauf achten, daß z.B. keine doppelten Merkmalsnamen und Fachschlüssel auftreten und einheitliche Bezeichnungen (z.B. Ausländer oder Nichtdeutsche) verwendet werden. Zur Vermeidung von Doppelarbeit bei der Erstellung der Metadaten hat das Bundesamt eine Dokumentation der bereits erstellten und in Bearbeitung befindlichen Metadaten zusammengestellt, die laufend aktualisiert wird und den Landesämtern auf dem zentralen Verbundserver StaNet zur Verfügung steht.

Bevor die beschlossenen Inhalte in GENESIS importiert und angeboten werden können, müssen die entsprechenden Wertedaten aus den jeweiligen Datenbeständen der Erhebung erstellt und in ein einheitliches Importformat umgesetzt werden. Die vorgegebenen Inhalte werden ausgewählt und die benötigten regionalen, sachlichen und ggf. zeitlichen Aggregate gebildet. Dazu wird eine Eigenentwicklung der Statistischen Ämter zur Auswertungsstatistischen Erhebungsmaterials genutzt: das „**Statistische Problemlösungsverfahren**“ kurz „SPLV“, das eigens für diese Aufgabe um eine GENESIS-Schnittstelle zur Datenquadererstellung erweitert wurde.

Bereits im Jahr 1996 hat der Arbeitskreis Informationstechnik „AKIT“ beschlossen, die Erstellung der GENESIS-Importdateien in den Arbeitsablauf der normalen Erhebungsaufbereitung zu integrieren. Die Länder Berlin, Hessen, Rheinland-Pfalz, Saarland, Sachsen-Anhalt und Thüringen haben sich im AKIT bereit erklärt, SPLV-Programme zu erstellen, die die in der „Unterarbeitsgruppe Datenbankinhalte“ beschlossenen und fachlich spezifizierten Inhalte im GENESIS-Importformat erzeugen.

### **Aus Alt mach Neu – GENESIS in Bayern**

Seit mehr als drei Jahrzehnten gibt es im Landesamt eine Statistische Datenbank: das **Bayerische Statistische Informationssystem** „BaSIS“. Die Menge der darin nachgewiesenen statistischen Zahlen hat eine stürmische Entwicklung genommen. Als die erste Version der Statistischen Datenbank am 13. November 1970 vom damaligen Bayerischen Staatsminister des Innern, Dr. Bruno Merk, in Betrieb genommen wurde, waren knapp 100 Daten je Gemeinde abrufbar. In den Folgejahren wurde der Datenbestand ständig ausgebaut. Die Aufnahme neuer Merkmale, aber auch die Fortschreibung des Datenbestandes durch die laufenden Statistiken trug zum stetigen Wachstum der Datenbank bei. Heute sind rund 16000 BaSIS-Merkmale und über 100000 Merkmals-/Zeitkombinationen verfügbar. Eine Übersicht über das Datenan-

gebot von BaSIS bietet der „Datenkatalog“. Er wird im Internet-Angebot des Landesamts [www.statistik.bayern.de](http://www.statistik.bayern.de) im Bereich „Daten und Fakten“ als PDF-Datei zum kostenlosen Download angeboten.

BaSIS kann Daten für unterschiedliche Gebietsstände speichern, was sich besonders in den 70er Jahren wegen der damaligen großen Kreis- und Gemeindegebietsreformen als zweckmäßig und notwendig erwiesen hat. Im Rahmen dieser Reformen wurde die Anzahl der kreisfreien Städte von 48 auf 25, die der Landkreise von 143 auf 71 und die der Gemeinden von 7004 auf nunmehr 2056 reduziert.

Mit diesem großen und vor allem gebietsstandsbereinigten Datenbestand ist die Statistische Datenbank zu einem unverzichtbaren Bestandteil in der Informationsversorgung für die Nutzer der amtlichen Statistik geworden. Sie liefert außerdem die Daten für einige Statistische Berichte sowie für die regional tief gegliederten Querschnittsveröffentlichungen des Landesamts: „Gemeindedaten“, „Kreisdaten“, „STATISTIK *kommunal*“ und die Kreisübersicht des Jahrbuchs.

Natürlich muß dieser wichtige Datenpool in GENESIS übernommen werden. Eine schwierige Aufgabe. Denn die Datenmodelle beider Datenbanken unterscheiden sich grundlegend. Bei BaSIS sind exakt drei Dimensionen fest vorgegeben: Raum, Sache und Zeit. Mehrere Sachdimensionen werden zu *einer* Dimension linearisiert.

Jedes in BaSIS nachgewiesene Merkmal hat eine eindeutige Merkmalsnummer. Sie beschreibt den sachlichen Inhalt des jeweils nachgewiesenen statistischen Werts. So steht z.B. das Merkmal mit der Nummer „BL1765“ für „Gestorbene Deutsche, weiblich, im Alter von 60 bis unter 65 Jahren“. Damit werden aber die *drei* Dimensionen „Geschlecht“, „Nationalität“ und „Altersgruppe“ auf *eine* sachliche Dimension reduziert und darüber hinaus zusammen mit der Benennung des Wertes („Gestorbene“) in *einer* Merkmalsnummer nachgewiesen.

Für den Abruf eines Wertes wird außerdem die sachliche und zeitliche Dimension zu *einer gemeinsamen* Angabe zusammengefaßt. Dazu wird in einem Datenbankabrufl die Merkmalsnummer für die Zeitauswahl um die sog. Zeitkomponente ergänzt: also z.B. „BL17659751“ wenn man die Jahressumme (Kennung „51“) des Kalenderjahres 1997 für das genannte Merkmal abrufen möchte.

Die Regionalachse ist in BaSIS auf die Regionaleinheiten „Gemeinden“, „Kreise“, „Regierungsbezirke“ und die „Landessumme“ festgelegt. Andere Regionalgliederungen sind nicht gespeichert, aber errechenbar.

Um nun diese Daten aus BaSIS in die mehrdimensionale Datenstruktur von GENESIS überführen zu können, wurde ein Umsetzprogramm erstellt. Es überführt mit entsprechenden Vorgaben jeden Wert aus dem Importbestand von BaSIS in einen Importsatz im GENESIS-Format. Dieser Weg wurde gewählt, um die Daten unabhängig vom IMS-System – dem BaSIS zu Grunde liegenden Datenbanksystem – konvertieren zu können.

Für die Umsetzung in das GENESIS-Importformat benötigt das Programm für jedes der rund 15000 Merkmale folgende Angaben:

- den Namen des Zieldatenquaders,
- die Position des Wertmerkmals im Zieldatenquader,

– die Koordinatenpositionen auf den regionalen, zeitlichen und sachlichen Achsen im Zieldatenquader.

Der aufwendigste Teil bei der Umsetzung in das GENESIS-Format ist die Rückführung der linearisierten Sachgliederungen von BaSIS in das mehrdimensionale System von GENESIS. Die Achsen der GENESIS-Datenquader bilden die Merkmale mit ihren Merkmalsausprägungen. Alle Merkmalsausprägungen haben in GENESIS einen eindeutigen Fachschlüssel. Für jedes Merkmal aus BaSIS muß dem Umsetzprogramm die Koordinatenposition in Form der GENESIS-Fachschlüssel mitgeteilt werden. Im genannten Beispiel „BL1765“ (siehe oben) sind dies „GESW“ für die Ausprägung „weiblich“, „NATD“ für „deutsch“ und schließlich „ALT060B65“ für „Altersgruppe 60 bis unter 65 Jahre“. Diese Koordinaten liegen auf den Achsen „Geschlecht“, „Nationalität“ bzw. „Altersgruppen“.

Häufig wurden für eine „saubere“ Datenstruktur in GENESIS Daten benötigt, die in BaSIS nicht physisch vorhanden waren, sondern berechnet werden mußten. So werden in BaSIS die Merkmale „Bevölkerung insgesamt“ und „Bevölkerung männlich“ gespeichert. Das Merkmal „Bevölkerung weiblich“ muß jedoch berechnet werden. Für eine komplette Achse im GENESIS-Quader „Bevölkerung nach Geschlecht“ werden die Merkmalsausprägungen „männlich“ und „weiblich“ benötigt. Die Werte für die Ausprägung „weiblich“ müssen folglich erzeugt werden. Auch dazu ist das Umsetzprogramm in der Lage, wenn man die entsprechende Formel definiert.

Auch die zeitliche Komponente wird durch das Umsetzprogramm in das GENESIS-Format umgesetzt. Hier sind je nach Format einige Zusatzangaben notwendig. Bei Stichtagsangaben, die in der „Zeitkomponente“ von BaSIS nicht abgebildet werden können, muß dem Umsetzprogramm eine komplette Liste der Zeitkomponenten und den entsprechenden Stichtagen zur Verfügung gestellt werden. Dies war beispielsweise bei den Wahlen der Fall.

Die regionale Komponente der Daten – also der Regionalschlüssel – ist in beiden Systemen grundsätzlich identisch und fließt als regionale Koordinate in die Datenquaderstruktur ein. Abbildung 7 zeigt die Vorlaufinformation, die für die Umsetzung eines Merkmals aus BaSIS in das GENESIS-Format notwendig ist, am Beispiel des Merkmals „BL1765“.

Vorlaufinformation			
BaSIS	GENESIS		
<b>BL1765</b>	<b>BEVSTF004#</b>	<b>01</b>	<b>GESW;NATD;ALT060B65</b>
Merkmalnr. in BaSIS			
Datenquadername	Nr. d. Wertmerkmals	Koordinate der Sachgliederung (Fachschlüssel der Ausprägung)	
GESW ..... Geschlecht „weiblich“			
NATD ..... Nationalität „deutsch“			
ALT060B65 ..... Altersgruppe „60 bis unter 65“			

▼  
**Umsetzprogramm**  
▼

Auszug aus der GENESIS Importdatei					
D;09161;GESW;NATD;ALT060B65;1983;21;e					
Ingolstadt			Alter 60 – 65		
weiblich		deutsch	Berichtsjahr	Wert	Qualität=endgültig

Abbildung 7

In einem zweiten Datenstrom wird dem Umsetzprogramm die Metadatenbeschreibung des umzusetzenden Datenquaders bereitgestellt. Sie wird vor den Zeilen der Wertedaten eingefügt und beschreibt die Eigenschaften des Datenquaders. So z. B. seinen „Namen“, die Erhebungszugehörigkeit, seine Achsen sowie seine Inhalte (Wertmerkmale) und deren Maßeinheiten. So weiß GENESIS, welche Daten importiert werden und wo sie im System hingehören.

Die Importsyntax, in der die Daten für GENESIS benötigt werden, ist das bereits beschriebene CSV-Format. Die „Datenzeilen“ enthalten die Importdaten. Die einleitenden „Kopfzeilen“ beschreiben deren Inhalte in den einzelnen Feldern.

#### Qualitätssicherung

So wurden bei der Migration des Datenbestandes aus BaSIS bislang gut 61 Millionen Daten konvertiert und nach GENESIS importiert. Im Rahmen der Qualitätssicherung (QS) mußte geprüft werden, ob die Daten auch richtig in GENESIS „angekommen“ sind. Hierfür wurde eine Microsoft Access-Anwendung entwickelt, die den Datenabgleich unterstützen und dokumentieren sollte.

In einem ersten Schritt wurden alle Daten aus BaSIS auf Landesebene in die Access-Datenbank importiert. In GENESIS wurde für jeden der knapp 500 Landesquader eine Abruftabelle mit allen Achsen erstellt, abgerufen, exportiert, auf den PC übertragen und ebenfalls in die Access-Datenbank importiert. Hierfür mußte diese GENESIS-Tabelle eine im QS-Konzept festgelegte Form haben.

Speziell für diese Prüfarbeiten wurden in GENESIS Merkmale und Referenzregeln definiert, die jeweils eine Landessumme aus Gemeinden, Kreisen und Regierungsbezirken bildeten. Diese mußten bei summierbaren Merkmalen mit den Daten des Landesquaders übereinstimmen. Damit wurde schon beim Import in die Access-Datenbank die Konsistenz der Summen aus den einzelnen hierarchischen Regionalquadern zu dem entsprechenden Landesquader geprüft.

Im nächsten Schritt wurde der Vergleich der Landesdaten aus GENESIS mit den Ursprungsdaten aus BaSIS halbautomatisch durchgeführt. Insgesamt waren hier ca. 80 000 Daten zu prüfen.

#### „Allgemeine Daten“ und „Zuordnungsdaten“

Auch die umfangreiche Sammlung von rund 100 in BaSIS gespeicherten „Allgemeinen Daten“ und „Zuordnungsdaten“ mußte in die neue Metadatenstruktur übergeführt werden. Neben Angaben wie Gemeindekoordinaten und Höhenlage, werden damit in BaSIS vorwiegend regionale Zuordnungen von Gemeinden bzw. Krei-

## Beispiel eines Dialogablaufs

①

Ausgehend vom Erhebungsverzeichnis sollen die Erhebungsmerkmale der Wanderungsstatistik angezeigt werden

②

Alle Datenquader mit dem Merkmal „Fortzüge über die Bundesgrenze“ sollen angezeigt werden

③

Der Regierungsbezirksquader wird an den Tabellenassistenten übergeben

④

Im ersten Schritt bietet der Tabellenassistent die Wertmerkmale des Datenquaders zur Selektion an

⑤

Im zweiten Schritt kann der Nutzer die gewünschten Achsen auswählen. Notwendige Angaben sind vorselektiert.

⑥

Im dritten Schritt erfolgt die Zeitauswahl

⑦

Aufgrund der Angaben im Dialog hat der Tabellenassistent eine Abruftabelle erstellt. Mit F11 erfolgt der Datenabruf

⑧

Die fertige Ergebnistabelle wird angezeigt.

sen zu anderen administrativen und nichtadministrativen Gebietseinheiten (z. B. Arbeitsamtsbezirke, Regionen, Fremdenverkehrsgebiete, Wahlkreise) sowie regionale Typisierungen und Kennzeichen (z. B. „Große Kreisstadt“, „Markt/Stadt“, „Einheits-/Mitgliedsgemeinde“, „Art der Fremdenverkehrsgemeinde“) abgebildet.

In GENESIS werden aus diesen Zuordnungsdaten Regionalmerkmale. Die einzelnen Regionaleinheiten sind die Ausprägungen des Merkmals. So gibt es beispielsweise ein Regionalmerkmal „REGION“ für die 18 bayerischen Regionen mit den Ausprägungen „REGION-01, Bayerischer Untermain (1)“, „REGION-02 Würzburg (2)“ ... „REGION-18, Südostoberbayern (18)“.

Außerdem muß es in GENESIS noch eine Referenzregel geben, die jeder Merkmalsausprägung des Ergebnismerkmals die Merkmalsausprägungen des Basismerkmals – bei Regionalmerkmalen im Normalfall Gemeinden – zuordnet. Diese Regel ist somit Dokumentation der Zugehörigkeiten und Bildungsvorschrift in einem. Abbildung 8 zeigt beispielhaft die Zuordnung von Gemeinden zu Regionen mittels einer Referenzregel.

Es wäre aufgrund der geringen Nachfrage sicher nicht sinnvoll, alle statistischen Daten für alle administrativen und nicht administrativen Regionalgliederungen physisch zu speichern, wenn sie aus den Gemeinde- oder Kreisergebnissen abgeleitet werden können. GENESIS nutzt hierzu das Instrument der Referenzregeln, das bei Bedarf und ohne Zutun des Benutzers die Werte mit den benötigten Achsen bildet.

Aber nicht nur die Daten müssen von BaSIS nach GENESIS übergeführt werden. Auch bestehende Daten-

bankabfragen aus BaSIS, die für regelmäßig wiederkehrende Datenlieferungen und Abonnements im Rahmen des Auskunftsdienstes erstellt wurden, müssen in die neue Abrufsyntax übersetzt werden. Der Lieferdatensatz wird sich damit für die Kunden allerdings ändern.

BaSIS ist eine reine Regionaldatenbank. In GENESIS dagegen gibt es kein solches primäres Ordnungskriterium. Damit bietet sich die Möglichkeit, das Datenangebot nach und nach um *Strukturdaten*, also sachlich tief gegliederte Daten auf hoher regionaler Ebene und *Zeitreihen* zu erweitern.

### Die weitere Entwicklung

Nach einer primär am funktionalen Ausbau orientierten Phase konzentrieren sich die Arbeiten nunmehr verstärkt auch auf Qualitätssicherung und Tuningmaßnahmen.

Dazu gehören die Verlässlichkeit und Stabilität des Systems, eine weitere Steigerung der Performance, die Konsolidierung der Dialoge und des Systemverhaltens – d.h. Beseitigung von eventuellen Inkonsistenzen in der Benutzerführung, in Masken und Menüs bzgl. Benennung und Dialogabläufen. Weiter aber auch die Umsetzung von Verbesserungsvorschlägen der Entwicklerteams, der GENESIS-Nutzergruppe und anderer GENESIS-Anwender.

Im Rahmen der funktionalen Vervollständigung wird es Kataloge für weitere GENESIS-Objekte geben, die Implementierung sog. Funktionalblöcke, also die Berechnung von Anteilswerten und Veränderungsrate. Weiter sollen bei der Aggregation und Ausgabe Qualitätsanzeiger berücksichtigt werden, wie z. B. „p“ (vorläufiges Ergebnis), „/“ (keine Angaben, da Zahlen nicht sicher genug) oder auch „x“ (Tabellenfach gesperrt, weil Aussage nicht sinnvoll).

Parallel hierzu wird vom Statistischen Bundesamt derzeit eine Client/Server-Version von GENESIS mit grafischer Benutzeroberfläche entwickelt.

Die noch anstehenden Aufgaben werden nach und nach entsprechend den zugeordneten Prioritäten und den zur Verfügung stehenden Kapazitäten abgearbeitet.

Dipl.-Volksw. Klaus Engelhardt  
Dipl.-Geogr. Irene Langer

Merkmalsausprägungen des Merkmals „Regionen“	
REGION-01	Bayerischer Untermain (1)
REGION-02	Würzburg (2)
REGION-03	Main-Rhön (3)
REGION-04	Oberfranken-West (4)
REGION-05	Oberfranken-Ost (5)
REGION-06	Oberpfalz-Nord (6)
:	:
Merkmalsausprägungen des Merkmals „Gemeinden“	
:	:
09377157	Waldershof, St
09377158	Waldsassen, St
09377159	Wiesau, M
09461000	Bamberg, Krfr.St
09462000	Bayreuth, Krfr.St
09463000	Coburg, Krfr.St
:	:

Referenzregel Regionen ↔ Gemeinden	
:	:
REGION-04	09461000
REGION-04	09463000
REGION-05	09377157
REGION-05	09462000
REGION-06	09377158
REGION-06	09377159
:	:

Abbildung 8

[1] Bayern in Zahlen, Ausgabe 11.1995, „Das gemeinsame neue statistische Informationssystem GENESIS“

1) Ein Mindestveröffentlichungsprogramm besteht aus einer Auswahl von Merkmalen, die hinsichtlich ihrer fachlichen, regionalen und zeitlichen Gliederung genau beschrieben sind sowie den entsprechenden Daten. Diese sind über die jeweiligen Medien zwingend zu veröffentlichen bzw. bereitzustellen und werden unter Angabe von Quellen und Veröffentlichungen einheitlich dokumentiert.

2) „Regionalstatistischer Datenkatalog des Bundes und der Länder“; Gemeinsamer Datenkatalog zur Bereitstellung regionalstatistischer Ergebnisse, der kurzfristig lieferbare Ergebnisse aus allen Bereichen der amtlichen Statistik auf Gemeinde- und Kreisebene umfaßt.