

Brecht, Beatrix

**Working Paper**

## Aufbau, Struktur und Anwendungen des sozio- ökonomischen Panels in INGRES

Diskussionsbeiträge - Serie II, No. 120

**Provided in Cooperation with:**

Department of Economics, University of Konstanz

*Suggested Citation:* Brecht, Beatrix (1990) : Aufbau, Struktur und Anwendungen des sozio-ökonomischen Panels in INGRES, Diskussionsbeiträge - Serie II, No. 120, Universität Konstanz, Sonderforschungsbereich 178 - Internationalisierung der Wirtschaft, Konstanz

This Version is available at:

<https://hdl.handle.net/10419/101547>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

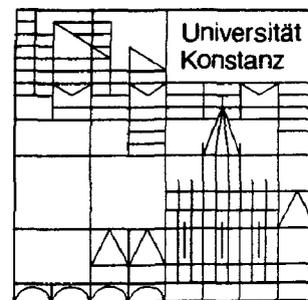
Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



**Sonderforschungsbereich 178**  
**„Internationalisierung der Wirtschaft“**

Diskussionsbeiträge

Juristische  
Fakultät

Fakultät für Wirtschafts-  
wissenschaften und Statistik

---

Beatrix Brecht

**Aufbau, Struktur und Anwendungen  
des sozio-ökonomischen Panels  
in INGRES**

**AUFBAU, STRUKTUR UND ANWENDUNGEN DES  
SOZIO-ÖKONOMISCHEN PANELS IN INGRES**

**Beatrix (Brecht**

Serie II - Nr. 120

November 1990

# AUFBAU, STRUKTUR UND ANWENDUNGEN DES SOZIO-ÖKONOMISCHEN PANELS IN INGRES

Beatrix Brecht

## **Zusammenfassung**

Dieser Beitrag schildert die Vorgehensweise, wie die Daten des Sozio-ökonomischen Panels, das vom Deutschen Institut für Wirtschaftsforschung (DIW) zur Verfügung gestellt wird, mit dem relationalen Datenbanksystem INGRES verwaltet werden können. Es wird auf das Importieren der Daten, Anwendungen und Bereitstellen von gewünschten Teildaten für statistische Softwarepakete sowie auf Aspekte der Datensicherheit eingegangen. INGRES stellt eine komfortable SQL-Version zur Verfügung, weshalb das Datenbanksystem gegenüber der SIR-Panel-Datenbank des DIW bevorzugt wurde.

# 1 Einführung

Das 1984 begonnene Sozio-ökonomische Panel (SOEP) umfaßt in seiner ersten Stichprobe rund 12.000 Personen in ca. 6000 Haushalten, wobei jährlich an die 1000 Variablen erhoben werden. Mit der nun sechsten Welle (1989) existiert eine Fülle von Informationen, die mit jeder neuen Panelwelle weiter zunimmt; daher ist es unverzichtbar, für die Datenhaltung und Auswertung ein System anzuwenden, das der wachsenden Komplexität gerecht wird. Um mit dieser umfangreichen Längsschnittstudie arbeiten zu können, wird ein Daten-Management benötigt, das in Form einer Datenbank einen einfachen und schnellen Zugriff auf die Informationen bietet. Das vom Deutschen Institut für Wirtschaftsforschung (DIW) verwendete relationale Datenbanksystem SIR (=Scientific Information Retrieval System) läßt sich im Vergleich zu einer modernen relationalen Datenbank wie z.B. INGRES (=Interactive Graphics and Retrieval System) wesentlich weniger komfortabel und effizient handhaben.

Im folgenden wird eine problemorientierte Darstellung für den Aufbau und die Strukturierung des SOEP mit INGRES gegeben. Ausgehend von dem Rohdatensatz wird das Einlagern der Daten, deren Strukturierung sowie das Abfragen und Auslagern der gewünschten Information geschildert. Das Ziel der Verwendung von INGRES im SFB-Projekt 178/C7 "Statistische Analyse von Gastarbeiterwanderungen" ist die Extraktion von Informationen über türkische Gastarbeiter, die in den Panelwellen des SOEP erfaßt sind. Diese Daten werden in einer Ergebnisdatei für weitere Analysen abgelegt, so daß mit gängigen Softwarepaketen wie SAS oder BMDP darauf zugegriffen werden kann. Die Darstellung soll als Hilfestellung für alle Nutzer des SOEP dienen, die nicht mit der SIR-Panel-Datenbank des DIW arbeiten, sondern eine relationale Datenbank wie ORACLE oder INGRES verwenden möchten, sowie als Grundlage für eine Diskussion über Verbesserungen dieser Benutzergruppe zur Verfügung stehen.

## 2 Funktionsweise von INGRES

Im Gegensatz zur relationalen SIR-Panel-Datenbank des DIW ist INGRES ein weiterentwickeltes relationales Datenbanksystem, das dem Anwender ein Höchstmaß an Flexibilität bietet. Bei diesem Datenbanktyp sind Daten, Zugriffslogik und Abspeicherungsmechanismus voneinander getrennt, womit eine einfache Handhabung und eine große Leistungsfähigkeit gewährleistet wird. Unter einer Datenbank versteht man im allgemeinen eine (computerisierte) Datensammlung in wohlgeordneten strukturierten Dateien. Zusammen mit dem Datenbank-Management-

System (DBMS), das als eine Gruppe von Funktionen zum Bearbeiten von Datenbanken – z.B. Abfragen, Einfügen und Manipulieren von Daten – definiert ist, erhält man ein Datenbanksystem (DBS).

Die ersten Datenbanksysteme waren hierarchisch aufgebaut, d.h. die Daten ließen sich als Baumstruktur in einem festen Verbund darstellen. Das weiterentwickelte vernetzte System verknüpft die Daten zu miteinander gekoppelten Sätzen, die sich überschneidende Datenmengen bilden. Die eingesetzten Sprachen in diesen Datenbanksystemen sind prozedural und verarbeiten nur jeweils einen Satz, wobei der Pfad zu diesem Satz explizit angegeben werden muß. Man hat also nicht nur mehrfache Anfragen, um einen bestimmten Satz zu finden, sondern man muß auch genau erfassen, wie die Daten abgespeichert sind.

In einem relationalen System dagegen liegen die Daten in Form von zweidimensionalen Tabellen als atomare Attribute vor. Es können gleichzeitig mehrere Sätze aus einer oder mehreren Tabellen abgerufen werden, wobei das relationale System automatisch zu den benötigten Daten führt, die wiederum in einer Relation ausgegeben werden. Eine relationale Sprache ist daher nichtprozedural; man teilt dem System lediglich mit, was zu tun ist, aber nicht, wie es vom System umzusetzen ist. Dadurch wird ein problemloser Zugriff auf die Daten ermöglicht, was bei der methodischen wie technischen Komplexität des Datenbestandes unverzichtbar ist. Edgar F. Codd hat in den 70er Jahren das theoretische Fundament zu diesem relationalen Datenbankmodell geschaffen, dessen Operationen mathematisch eindeutig definiert sind. Er formulierte 1985 zwölf Regeln, denen ein relationales Datenbanksystem genügen muß, wobei die wichtigste Grundregel fordert, daß alle Daten ohne Ausnahme mit relationalen, d.h. mengenorientierten Operationen bearbeitet werden können, ohne daß irgendwelche Sonderwerkzeuge außerhalb des relationalen Modells benötigt werden. Eine relationale Datenbank hat desweiteren keine Verknüpfungsbeschränkungen, weder Beschränkungen bei der Teilung einer Relation (Projektion, Restriktion), noch beim Zusammenfügen (JOIN). Es gibt keine Zeiger und keine Baumstrukturen.

1973 wurde ein "University-INGRES" in Berkeley entwickelt, das in erweiterter Version 1980 von Relational Technology Inc. auf den Markt gebracht wurde und mittlerweile neben ORACLE und DB2 zu den am weitesten verbreiteten relationalen Datenbanksystemen gehört. Außer dem Vorteil, daß es ein echtes relationales Datenbanksystem ist, enthält INGRES alle DBMS-Standardfunktionen wie Transaktion oder Mehrfachzugriff, eine Performance (Ablaufoptimierung), teilbare Prozesse, eigenes Maskengenerieren mit VIFRED (= Visual Forms Editor) sowie portierbare Applikationen durch 4GL und ist verfügbar sowohl auf Großrechneranlagen

als auch auf PC<sup>1</sup>. Der zentrale Punkt ist jedoch die SQL-Schnittstelle.

SQL - Structured Query Language - geht weit über eine einfache Abfragesprache hinaus und ist zum Standard für Abfragesprachen relationaler Datenbanken erhoben worden, d.h. es wird darauf hingearbeitet, SQL für jedes beliebige DBMS einzusetzen bzw. die Einbettung über "Gateways" zu ermöglichen. Noch allerdings gibt es in der Praxis große Schwierigkeiten, auch wenn mittlerweile in vorhandene Systeme wie dBaseIV SQL eingeführt wurde. Dabei bietet SQL nicht nur den Vorteil einer nichtprozeduralen Programmiersprache, sondern auch die Möglichkeit, in die traditionellen Programmiersprachen wie Pascal, Cobol, Fortran oder C - in dem INGRES übrigens programmiert wurde - eingebettet zu werden, so daß bereits vorhandene Anwendungsprogramme miteingebunden werden können.

Durch die SQL-Operationen lassen sich in den spaltenorientierten relationalen Datenbanksystemen beliebige Beziehungen über die Elemente der Satzarten bereitstellen, wobei die Verknüpfungen grundsätzlich über sogenannte Schlüsselspalten stattfinden. Dabei muß eine relationale Datenstruktur garantieren, daß jedes Datenelement in den Relationen (=Tabellen) nur einmal oder als streng kontrollierte Wiederholung vorhanden ist und die Daten voneinander unabhängig sind. Daraus folgt, daß man beim Datenbankdesign zunächst die Strukturierung von Objekten sowie die Festlegung von Objektbeziehungen, z.B. in einem Entity-Relationship-Modell, vornehmen muß, um eindeutige Datenstrukturen zu erzeugen, d.h. es darf keinen Datensatz mit identischen Spaltenkombinationen geben. Dieses logische und konzeptionelle Design wird dann in Relationen umgesetzt, wobei der Primärschlüssel festgelegt wird, der die Eindeutigkeit eines Datensatzes garantiert. Durch die Normalisierung schließlich, d.h. das Auflösen nach funktionalen Abhängigkeiten, wird zusätzlich eine redundanzarme Speicherung der Daten bewirkt.

SQL läßt sich in die folgenden Funktionsbereiche gliedern:

- Datendefinition (Data Definition Language, DDL): beinhaltet z.B. das Anlegen und Löschen von Tabellen und Datensichten (VIEWS) [create, drop]
- Datenmanipulation (Data Manipulation Language, DML): Einfügen, Modifizieren und Löschen von Daten [insert, update, delete]
- Queries: Abfragen der Information mit Hilfe des Befehls select

---

<sup>1</sup>Voraussetzungen für die PC-Installation: 100% IBM-kompatibel, MS-DOS-Version  $\geq 2.0$  und Arbeitsspeicher  $\geq 640$  kB.

- Datenkontrolle, Benutzerautorisation (Data Control Language, DCL): Vergabe von Zugriffsbeschränkungen in einem Mehrbenutzersystem für die Datensicherheit [grant], Transaktionsregeln [begin, commit, rollback]

Desweiteren kann SQL als Data Dictionary fungieren, d.h. es stellt Informationen über alle definierten Objekte (Tabellen, VIEWS, Zugriffsrechte) zur Verfügung, wodurch die Dokumentation der Datenbank erleichtert wird.

Vor allem die komfortable Handhabung des SQL läßt INGRES gegenüber dem SIR-Datenbanksystem als überlegen gelten; so ist die Syntax beispielsweise einer einfachen Abfrage

```
SELECT Spalte(n)
      FROM Tabelle;                                (1)
```

auch ohne zusätzliche Kenntnisse in SQL sofort einsichtig.

Nach diesem Überblick über INGRES und SQL wird nun konkret auf die Anwendungen für das SOEP eingegangen. Dabei müssen Grundkenntnisse des interessierten Lesers in SQL sowie im Umgang mit dem Datenbanksystem INGRES vorausgesetzt werden, die relativ einfach erarbeitet werden können. Als einführende Literatur sei auf Date (1987), RTI (1987), van der Lans (1988) oder Berger (1989) hingewiesen.

### 3 Das Sozio-ökonomische Panel (SOEP)

Das jährlich erhobene SOEP<sup>2</sup> stellt eine repräsentative Längsschnittuntersuchung über Fragen wie Einkommen, Erwerbsbeteiligung, Wohnsituation, Mobilität sowie subjektive Einschätzungen dar. Die erste 1984 durchgeführte Stichprobe umfaßte 12.245 Personen in 5.921 Haushalten; in der 5. Welle 1988 konnten davon noch 10.023 Personen und 4.814 Haushalte erreicht werden. Die Grundgesamtheit der Stichproben besteht aus der wohnberechtigten Bevölkerung Westdeutschlands (einschließlich dem westlichen Teil Berlins) und ist in zwei Teilstichproben getrennt. Teilstichprobe A ist definiert als die (west-)deutsche Wohnbevölkerung einschließlich sonstiger Ausländer, Teilstichprobe B in Haushaltsvorstände mit griechischer, italienischer, jugoslawischer, spanischer oder türkischer Staatsangehörigkeit. Mittlerweile steht die 6. Welle (1989) zur Auswertung zur Verfügung, die 7. Welle ist bereits in der Feldphase abgeschlossen.

---

<sup>2</sup>zur Nutzungsmöglichkeit wende man sich direkt an das DIW, Berlin.

Die Entwicklung der Fallzahlen in einem Panel weist auf einen kontinuierlichen Ausfall von Personen bzw. Haushalten hin; dabei ist die Hauptursache neben dem Wegzug ins Ausland bzw. unbekannter Zielort eindeutig die Verweigerung an einer erneuten Teilnahme. Dennoch sind rund 70% der erstmalig befragten Personen von 1984 noch in der 5. Welle erhoben, wobei allerdings auch Zugänge in Form von Geburten, Zuzug in einen erfaßten Haushalt oder Erreichen des Befragungsalters (16 Jahre) enthalten sind sowie Neugründungen von Haushalten durch Stammpersonen alter Haushalte (vgl. Tabelle 1).

Tabelle 1: Entwicklung der Panelstichprobe, Welle 1 - 5  
(Fallzahlen netto, d.h. Haushalte mit realisierten Interviews)

	1984	1985	1986	1987	1988
	Welle 1	Welle 2	Welle 3	Welle 4	Welle 5
<u>Stichprobe A:</u>					
<u>Deutsche</u>					
Haushalte	4 582	4 141	3 962	3 910	3 743
Haushaltsmitglieder	11 406	10 588	10 128	9 925	9 468
Befragungspersonen	9 114	8 520	8 200	8 042	7 692
Personeninterviews	9 076	8 372	8 009	7 868	7 481
<u>Stichprobe B:</u>					
<u>Ausländer</u>					
Haushalte	1 393	1 181	1 128	1 116	1 071
Haushaltsmitglieder	4 805	4 198	3 993	3 959	3 756
Befragungspersonen	3 181	2 813	2 712	2 716	2 626
Personeninterviews	3 169	2 718	2 637	2 648	2 542
<u>Gesamt:</u>					
Haushalte	5 921	5 322	5 090	5 026	4 814
Haushaltsmitglieder	16 211	14 786	14 121	13 884	13 224
Befragungspersonen	12 295	11 333	10 912	10 758	10 318
Personeninterviews	12 245	11 090	10 646	10 516	10 023

Quelle: Infratest Sozialforschung 1989, S.44.

Die Erhebung erfolgt in der Regel durch persönliche Befragung eines Interviewers; nur in Ausnahmefällen wird ein schriftliches Interview durchgeführt. Dem Interviewer stehen als Erhebungsinstrumente ein Haushaltsfragebogen mit Fragen zur Haushaltszusammensetzung, -einkommen

und zur Wohnung zur Verfügung, ein Personenfragebogen, der die Bereiche Einkommen, Erwerbstätigkeit, Freizeit, Gesundheit, Politik, Demographie, Zufriedenheit und Angaben zur Person abdeckt, sowie ein Adressenprotokoll zur Dokumentation der Befragung bzw. Angaben über eine Antwortverweigerung, das vom Interviewer ausgefüllt wird. Für die Teilstichprobe B liegen der Haushalts- und Personenfragebogen zweisprachig in der jeweiligen Landessprache mit deutschen Untertexten vor. Jede Welle beinhaltet neben den jährlich erhobenen Fragen einen besonderen Themenschwerpunkt - in der 5. Welle z.B. eine Vermögensbilanz - , der als zusätzliche Hintergrundinformation zur Verfügung steht. Einen Überblick über die Schwerpunkte und Analysemöglichkeiten des SOEP gibt Hanefeld (1987).

## 4 Der Datensatz des SOEP

Die einzelnen Panelwellen lassen sich in jeweils 14 (für die Wellen 1-4), 16 (für die Welle 5) bzw. 18 (für die 6. Welle) Einzeldateien gliedern, in denen die Fragen vercodet abgelegt sind. Dabei unterteilen sich die Dateien jeweils in eine Datendatei (erkennbar am Kürzel 'DAT') und in eine Dictionarydatei ('DIC'), die eine Dokumentation zur Datendatei mit Variablennamen, Ausprägungen und Format enthält. Es ist empfehlenswert, sich diese Dictionarydateien ausdrucken zu lassen, da sie als Arbeitsgrundlage neben dem Benutzerhandbuch des DIW unerlässlich sind.

Prinzipiell kann man zwei Typen von Dateien unterscheiden: die personenbezogenen Tabellen (erkennbar am 'P' in der 5. Stelle des Dateinamens) und die haushaltsbezogenen Tabellen (entsprechend ein 'H'). Ausnahmen von dieser Systematik bilden die Zusatzerhebungen und die kinderbezogenen Dateien ('KIND').

Der vierte Buchstabe des Dateinamens kennzeichnet die Welle, ein 'A' steht für die 1. Welle (1984), 'B' für die 2. Welle etc. Schließlich enthält die jeweils letzte Panelwelle aktualisierte Hochrechnungs- und Gewichtungdaten ('PHRF' und 'HHRF'), sowie die Dateien PPFAD und HPFAD, die den Gesamt-Stichprobenbestand und den Befragungsstatus in den jeweiligen Wellen dokumentieren. D.h. in den Dateien PPFAD bzw. HPFAD sind alle Personen bzw. Haushalte enthalten, die in mindestens einem DATxPBRU bzw. DATxHBRU abgespeichert sind.

Tabelle 2: Übersicht der Datendateien des Sozio-ökonomischen Panels (Welle 1-6)

Übergreifende Dateien		Welle 1	Welle 2	Welle 3	Welle 4	Welle 5	Welle 6
Name	DATPPFAD	DATAH	DATBH	DATCH	DATDH	DATEH	DATFH
(Variablenzahl/Fallzahl)	(13/18692)	(93/5921)	(140/5322)	(143/5090)	(142/5026)	(129/4814)	(118/4690)
	DATHPFAD	DATAKIND	DATBHBRUTTO	DATCHBRUTTO	DATDHBRUTTO	DATEHBRUTTO	DATFHBRUTTO
	(13/7015)	(27/3924)	(39/6142)	(40/5878)	(40/5512)	(42/5403)	(40/5213)
	DATPHRF	DATAP	DATBKIND	DATCKIND	DATDKIND	DATEKIND	DATFKIND
	(28/18692)	(276/12290)	(25/3450)	(25/3201)	(29/3050)	(20/2849)	(20/2734)
	DATHHRF	DATAPPAUSL	DATBP	DATCP	DATDP	DATEP	DATFP
	(28/7015)	(46/3175)	(443/11090)	(404/10646)	(392/10516)	(359/10023)	(462/9710)
		DATAPBIO	DATBPAUSL	DATCPAUSL	DATDPAUSL	DATEPAUSL	DATFPAUSL
		(484/12290)	(68/2624)	(77/2513)	(77/2517)	(73/2421)	(74/2374)
		DATAPBRUTTO	DATBPBRUTTO	DATCPBRUTTO	DATDPBRUTTO	DATEPBRUTTO	DATFPBRUTTO
		(12/16252)	(29/16977)	(32/16536)	(32/15984)	(32/15931)	(31/15619)
		DATAPKAL	DATBPKAL	DATCPKAL	DATDPKAL	DATEPKAL	DATFPKAL
		(348/12290)	(271/11090)	(273/10646)	(278/10516)	(274/10023)	(270/9710)
						DATEV	DATARTK
						(121/4564)	(9/33496)
							DATEINK
							(9/27981)

Die einzelnen Dateien korrespondieren zu den entsprechenden Fragebögen, so stehen z.B. in DATEPKAL die Variablen des Kalendarium aus dem Personenfragebogen der 5. Welle, in DATBPAUS sind die Fragen des Personenfragebogens für Ausländer der 2. Welle vercodet. Eine besondere Bedeutung kommt den Dateien DATxP bzw. DATxH und DATxPBRU bzw. DATxHBRU zu. In den letzteren sind die Basisinformationen der jeweiligen Panelwelle enthalten, also auch bei nicht realisierten Interviews. Dagegen beinhalten DATxP bzw. DATxH lediglich die in der jeweiligen Welle erfolgreich befragten Personen bzw. Haushalte. Somit lassen sich aus diesen Dateien Rückschlüsse ziehen, wie umfangreich der Stichprobenbestand ist, der an den Panelwellen teilgenommen hat bzw. wann – und unter welchen Umständen – ein Haushalt oder eine Person aus dem Panel ausgeschieden ist.

Die Variablennamen schließlich, die in den Einzeldateien enthalten sind, orientieren sich in der Regel sehr eng an dem Fragebogen. So kennzeichnet der 1. Buchstabe wieder die Panelwelle, der 2. Buchstabe weist auf personenbezogene bzw. haushaltsbezogene Informationen hin; schließlich folgen die Fragennummern aus dem Original-Fragebogen, gegebenenfalls weiter nach Item gegliedert bzw. mit einem Zusatz (A bei ausländerspezifischen Fragen, B bei erstmaliger Befragung, G für wiederholte Befragung) zur weiteren Identifikation versehen. Beispielsweise bedeutet CP82A01, daß hier die Ausprägungen aus dem Personenfragebogen für Ausländer, 3. Welle, der Frage 82 mit Item 1 abgelegt sind.

Die gesamte Information des SOEP ist somit in eine Vielzahl von Tabellen zerlegt. INGRES ermöglicht die Datenhaltung entsprechend dieser Tabellen, wobei die Relationen von rechteckiger Datenstruktur sind, d.h. alle abgespeicherten Datensätze einer Tabelle besitzen gleichviele Variablen. Außer den Daten selbst müssen auch deren Format und Beschreibung (Label) erfaßt werden.

Um Abfragen aus mehreren Relationen gleichzeitig realisieren zu können, ist ein ein-eindeutiger Schlüssel für jeden Datensatz notwendig. Die eindeutige Identifikation (Primärschlüssel) eines jeden Haushalts wird durch die Variable HHNR bzw. HHNRAKT als weiteres Identifikationskennzeichen, falls der Haushalt durch Abspaltung – d.h. Umzug von Personen aus bisherigen Panel-Haushalten und Gründung eines eigenen neuen Haushalts – entstanden ist, gewährleistet.

Für personenbezogene Untersuchungen steht die Variable PERSNR als eindeutiger Identifikator zur Verfügung, der für jede Person unveränderlich in sämtlichen Wellen erhalten bleibt. Dadurch kann auch der Haushaltswechsel einer Person problemlos nachvollzogen werden. Beim

praktischen Umsetzen ist vor allem darauf zu achten, daß diese Identifikatoren in jeder Relation enthalten sind.

## 5 Das Anlegen der INGRES-Panel-Datenbank

Obwohl für das konkrete Analyseziel – Aussagen über die Rückkehrbereitschaft sowie Einflüsse auf die Aufenthaltsdauer türkischer Gastarbeiter in Westdeutschland – ein weniger umfangreicher Teil des SOEP als Informationsgrundlage ausreichend gewesen wäre, wurde im Hinblick auf weitere Projektarbeiten das vollständige SOEP in die INGRES-Datenbank importiert<sup>3</sup>.

Die folgenden Ausführungen beziehen sich auf die praktische Handhabung der Daten, wobei aufgrund der Datenmenge an einer COMPAREX-Großrechneranlage, System VM/CMS, der Universität Konstanz gearbeitet wurde; implementiert ist die INGRES-Version 5.0.

Der erste Schritt besteht im Einlesen der Panel-Rohdaten vom Band des DIW auf die Rechenanlage. Die vom DIW mitgelieferte Bandsatzbeschreibung enthält alle relevanten Angaben, insbesondere ist die Label-Kennzeichnung wichtig. Vom aufgespannten Band lassen sich die Dateien über '*FILEDEF OUT/FILEDEF IN*'-Kommandos auf eine Platte der User-ID einlesen. Idealerweise sollte man sich dazu eine temporäre Platte mit ausreichender Kapazität anlegen, auf die man die Daten vom Band zwischenspeichert, bis sie in die INGRES-Datenbank übernommen werden. Es ist desweiteren empfehlenswert, die Dictionary-Dateien ausdrucken zu lassen und sie anschließend von der temporären Platte zu löschen.

Als nächster Schritt soll die INGRES-Datenbank am Rechner eingerichtet werden. Voraussetzung dafür ist zunächst die allgemeine INGRES-Zugangsberechtigung, die nur der Datenbankadministrator erteilen kann. Schließlich muß man an die Platte mit INGRES-Software angeschlossen sein (*LINKTO INGRES*). Das Einrichten einer völlig neuen Datenbank erfolgt dann durch den Befehl '*CREATEDB datenbankname*'. Über das Kommando '*INGRES datenbankname*' gelangt man in das INGRES-Menü, aus dem der Punkt '*TABLES*' mit dem folgenden Unterpunkt '*CREATE*' auszuwählen ist. Die neuen Tabellennamen wurden bewußt entsprechend dem Panel-Dateienaufbau gewählt, um die Systematik beibehalten zu können. Nach der Festlegung des Tabellennamens werden die jeweiligen Spaltennamen sowie das Datenformat eingegeben. Für diese Arbeit benötigt man die ausgedruckten Dictionary-Dateien, in denen unter 'Data

---

<sup>3</sup>Der besondere Dank für ihre Unterstützung gilt Frau Dipl. Math. Hanne Schröder, Universität Konstanz.

List' die Variablennamen und unter 'Format' die in der Reihenfolge abgelegten Formate aufgelistet sind. Es empfiehlt sich, unbedingt über das 'c'-Format einzulesen, da die Ausprägungen nicht nur als Integerzahlen vorkommen, sondern auch Buchstaben – z.B. 'M' als Missing Value Code – enthalten sein können.

Die Eingabemaske für die Datei DATBPBRU sieht z.B. wie folgt aus (Abbildung 1):

Abbildung 1

INGRES TABLE UTILITY Database: panel  
 Creating a New Table

Enter the name of the new table: datbpbru

Enter the column specifications of the new table:

Column Name	Data Format
hhnr	c5
persnr	c6
hhnrakt	c5
welle	c2
bpnrakt	c2
bgeburt	c4
bsex	c1
bstell	c2
bpnat	c1
bstistat	c1

Die fertige Tabelle wird schließlich mit 'SAVE' abgespeichert.

Für das Erstellen weiterer neuer Tabellen läßt sich der Vorgang beschleunigen, indem nach Eingabe des neuen Tabellennamens eine bereits existierende Tabelle geladen wird (Menüpunkt 'GetTableDef'), so daß in dieser lediglich Korrekturen vorgenommen werden müssen. Insbesondere bei ähnlichen Variablennamen, wie z.B. in Tabellen DATAKIND und DATBKIND, kann dieser Kniff sehr zeitsparend sein.

Auf zwei wichtige INGRES-Formalitäten und den daraus folgenden Konsequenzen soll noch hingewiesen werden:

- Formal kann eine Variablenbezeichnung in INGRES maximal 12 Buchstaben besitzen, wovon die ersten 9 eindeutig sein müssen. Ratsam ist es allerdings, lediglich einen Variablennamen mit 8 Buchstaben zu wählen, um beim späteren Arbeiten mit Abfragen die Zwischenergebnisse auf die Arbeitsplatte am Großrechner zu speichern, wobei die übliche 8-Buchstaben-Konvention gilt.
- Ein Problem ergibt sich aus der Tatsache, daß die maximale Spaltenanzahl einer Tabelle bei 127 Spalten liegt, d.h. größere Paneldateien, wie z.B. DATAPBIO oder die DATxP-Dateien, müssen in mehrere Dateien gesplittet werden. Dabei ist darauf zu achten, daß wichtige Informationsvariablen, wie die HHNR, HHNRAKT und die PERSNR, in jeder Teildatei vorhanden sind.

Als Beispiel wird die Datei DATAPKAL angeführt; sie kann in die folgenden Teildateien gesplittet werden:

DATAPKA1 beginnt mit den Variablen HHNR, HHNRAKT, PERSNR etc. und endet mit der Variablen AP36108. Ihre Variablenanzahl beträgt somit 114.

DATAPKA2 benötigt wieder den 'Kopf' mit HHNR, HHNRAKT, PERSNR, WELLE, SAMPEL, LEERSATZ und geht dann von AP3701 bis AP37118. Variablenanzahl: 124.

DATAPKA3 schließlich erhält ebenfalls den 'Kopf' von HHNR bis LEERSATZ sowie die restlichen Variablen ab AP37119 bis APERV30. Das ergibt 123 Variablen.

Sind die Tabellenstrukturen erstellt, werden im nächsten Arbeitsschritt die Daten in die INGRES-Datenbank importiert. Dieser Vorgang läßt sich sehr komfortabel mit SQL lösen. Die dazugehörige Befehlsstruktur lautet:

```
COPY TABLE tabellenname (spaltenname 1 = c format,  
                           spaltenname 2 = c format,  
                           :  
                           spaltenname n = c0nl) (2)  
FROM 'fm : dateiname.ft'
```

(INGRES benutzt Dateinamen nach der PC-Konvention, d.h. 'Platte (Laufwerk): Name.Extention'; in VM/CMS wird für Dateien die Konvention 'Name\_Extention\_Platte' angewandt.)

Um sich die Arbeit, nochmals sämtliche Variablennamen mit ihren Datenformaten eingeben zu müssen, zu ersparen, bedient man sich praktischerweise folgenden Verfahrens:

Im SQL-Monitor listet man über den Befehl '*HELP tabellenname*' die angelegte Tabelle mit den Formaten auf und speichert sie unter einem Dummynamen in einer Datei ab (z.B. '*tabellenname.copy*'). Nachdem man in den XEDIT vom SQL-Monitor gewechselt hat, läßt sich diese Datei mit dem Kommando '*GET tabellenname copy*' aufrufen und bearbeiten. Um die geforderte Befehlsstruktur (s. oben) zu erreichen, müssen die Informationstexte aus der durch den Help-Aufruf erzeugten Datei entfernt werden und stattdessen die SQL-Syntax (2) ergänzt werden. Dabei steht 'fm' für den Filemode, also die Plattenangabe, wo die Daten zu finden sind, 'dateiname' ist die korrespondierende Datendatei zur Tabelle ('*tabellenname*') und 'ft' steht für den Filetype der Datendatei. Die letzte einzulesende Variable erhält die Formatangabe 'c0nl', womit bewirkt wird, daß die Zeile fertig eingelesen und zu einer neuen Zeile gesprungen wird ('nl' = new line).

Das Zuordnen der Formate und Einsetzen der Kommata am Zeilenende läßt sich im XEDIT problemlos durch die Befehlsfolge

$$SET\_ARBCCHAR\_ON\$ \quad (3)$$
$$c\_c\$zahl / =\_czahl / \* \*$$

erreichen, wobei für 'zahl' nacheinander die vorkommenden Formatwerte eingesetzt werden.

Hat die Datei schließlich die Syntax gemäß (2), wird sie mit dem Kommando '*SAVE tabellenname ft*' für weitere Verwendungen abgespeichert, und man wechselt mit dem Kommando '*FILE*' in den SQL-Monitor über. Mit '*GO*' wird der Datenimport gestartet; den erfolgreichen Ablauf kann man sich über '*SELECT \* FROM tabellenname*' auf dem Bildschirm anschauen.

Bei gesplitteten Dateien ist das Importieren der Daten etwas komplizierter, da die zusätzlich vorhandenen Variablennamen im 'Kopf' jeder Teildatei ebenfalls den richtigen Daten zugewiesen werden müssen.

Indem man im SQL-Monitor über '*HELP tabellenname*' jede einzelne Tabelle der Reihe nach aufruft und unter demselben Namen abspeichert, erzeugt man eine zusammenhängende "große" Formatdatei. Im XEDIT werden dann wieder die Informationstexte sowie der mehrfache Variablen'kopf' aus dieser Gesamtdatei entfernt und schließlich die SQL-Syntax (2) eingefügt. Aus

dieser Gesamtdatei muß für jede Teildatei entsprechend der angelegten Dateimaske eine eigene Importdatei erstellt werden. Dazu speichert man die Gesamtdatei zunächst ab, um sie für jede Teildatei als Grundlage verwenden zu können. Das Erstellen der Importdatei erfolgt unter Zuhilfenahme der sogenannten 'd'-Formate, die es ermöglichen, Variablen zu überlesen. Dazu werden die nicht einzulesenden Spalten abgezählt und ihre Formate addiert. Möglichst viele aufeinanderfolgende, nicht einzulesenden Spalten werden in einer Dummyvariablen (z.B. X1) zusammengefaßt:

$$X1 = D \text{ anzahl}, \quad (4)$$

wobei 'anzahl' maximal den Wert 255 erreichen darf. Für weitere nicht relevante Spalten kreiert man entsprechend zusätzliche Dummyvariablen X2, X3, ....

Für das vorherige Beispiel der in Einzeldateien gesplitteten Datei 'DATAPKAL' sehen die Importdateien gemäß Abbildung 2 aus.

Nachdem somit die Daten des Sozio-ökonomischen Panels in die INGRES-Datenbank importiert wurden, ist die Panel-Datenbank einsatzbereit. Im folgenden werden einige Hinweise für das Strukturieren, für die Queries und das Auslagern von erstellten Datensätzen, um mit anderen Software-Paketen darauf zugreifen zu können, gegeben.

Abbildung 2: Importieren der Daten bei gesplitteter Datei

```

COPY TABLE datapka1 (hhnr=c5, COPY TABLE datapka2 (hhnr=c5, COPY TABLE datapka3 (hhnr=c5,
    hhnrakt=c5, hhnrakt=c5, hhnrakt=c5,
    persnr=c6, persnr=c6, persnr=c6,
    welle=c2, welle=c2, welle=c2,
    sampel=c1, sampel=c1, sampel=c1,
    leersatz=c1, leersatz=c1, leersatz=c1
    ap3601=c1, x1=d108, x1=d108,
    ap36107=c1, ap3701=c1, x2=d118,
    ap36108=c1, ap3702=c1, ap37119=c1,
    x1 = d118, ap37118=c1,
    x2 = d154, x2=d154,
    x3 = d0n1) x3=d0n1)
    aperv29=c1,
    aperv30=c0n1)
FROM 'b:datapka1.daten' FROM 'b:datapka1.daten' FROM 'b:datapka1.daten'

```

## 6 Anwendungen in INGRES

Das Hauptziel der Anwendungen von Datenbanken besteht darin, möglichst komfortabel die in den Tabellen abgelegten Informationen abzufragen. In SQL geschieht dies mit Queries (Abfragen), die durch den Befehl 'SELECT' ausgeführt werden. Das einfachste Syntax ist durch

```
SELECT * FROM tabellename; (5)
```

gegeben, wobei das Asteriskzeichen alle Spalten der Tabelle in der Abfrage miteinbezieht. Konditionen, um das Ergebnis einer Query einzugrenzen, werden durch die Anweisung 'WHERE' gegeben. Mit

```
SELECT hhnr FROM datbpaus WHERE sampel = "2";
```

erhält man die Haushaltsnummern (hhnr) aller türkischer Haushalte (sampel = 2 steht für die türkische Nationalität) der 2. Welle aus der Datei datbpaus<sup>4</sup>.

Der Vorgang (5), eine oder mehrere Spalte(n) aus einer Tabelle herauszugreifen, bezeichnet man als Projektion. Werden Bedingungen über das Befehlswort 'WHERE' angegeben, wird eine Restriktion angewendet. Die Bedingungen können aus den bekannten Vergleichsoperatoren (=, >, <, >=, <=, <>) bestehen, aus logischen Operatoren (AND, OR, NOT) und/oder aus sogenannten Subqueries (IN, EXIST), die einem zusätzlichen (Inner-)Select entsprechen.

Eine weitere wichtige Anwendungsmöglichkeit ist der JOIN-Operator, durch den mehrere Tabellen miteinander verknüpft werden. Dazu benötigen die zu verknüpfenden Tabellen eine gemeinsame Spalte, die es erlaubt, die Zeilen sinnvoll miteinander zu verbinden. Da in der Paneldatenbank jede Tabelle die Variable HHNR bzw. PERSNR enthält, können die einzelnen Dateien problemlos miteinander verknüpft werden. Die Syntax dazu sieht folgendermaßen aus:

```
SELECT a.spalte1, a.spalte2, b.spalte3, b.spalte4, usw.  
FROM a, b (6)  
WHERE a.spalte1 = b.spalte1;
```

Es können beliebig viele Spalten aus den Tabellen a und b herausgegriffen werden; im Beispiel (6) ist unterstellt, daß die 'spalte1' in beiden Tabellen vorhanden ist (setze z.B. spalte1 = 'hhnr')

---

<sup>4</sup>Leider können an dieser Stelle die vielen Funktionen und Möglichkeiten von INGRES nur gestreift werden, um eine Anregung zu geben. Zur Vertiefung und Weiterbildung muß auf die angegebene Literatur verwiesen werden. Im Rahmen des Projektes ist geplant, ein INGRES-Tutorial, speziell zugeschnitten auf die Anwendungen mit dem Sozio-ökonomischen Panel, zu erstellen.

für die Panel-Datenbank). Die Bedingungen im 'Where'-Satz können entsprechend kompliziert gestaltet werden. In einem relationalen Datenbanksystem lassen sich somit fast beliebig viele Tabellen miteinander verbinden, wobei allerdings durch eine zunehmende Anzahl auch die (logische) Komplexität der Abfrage steigt.

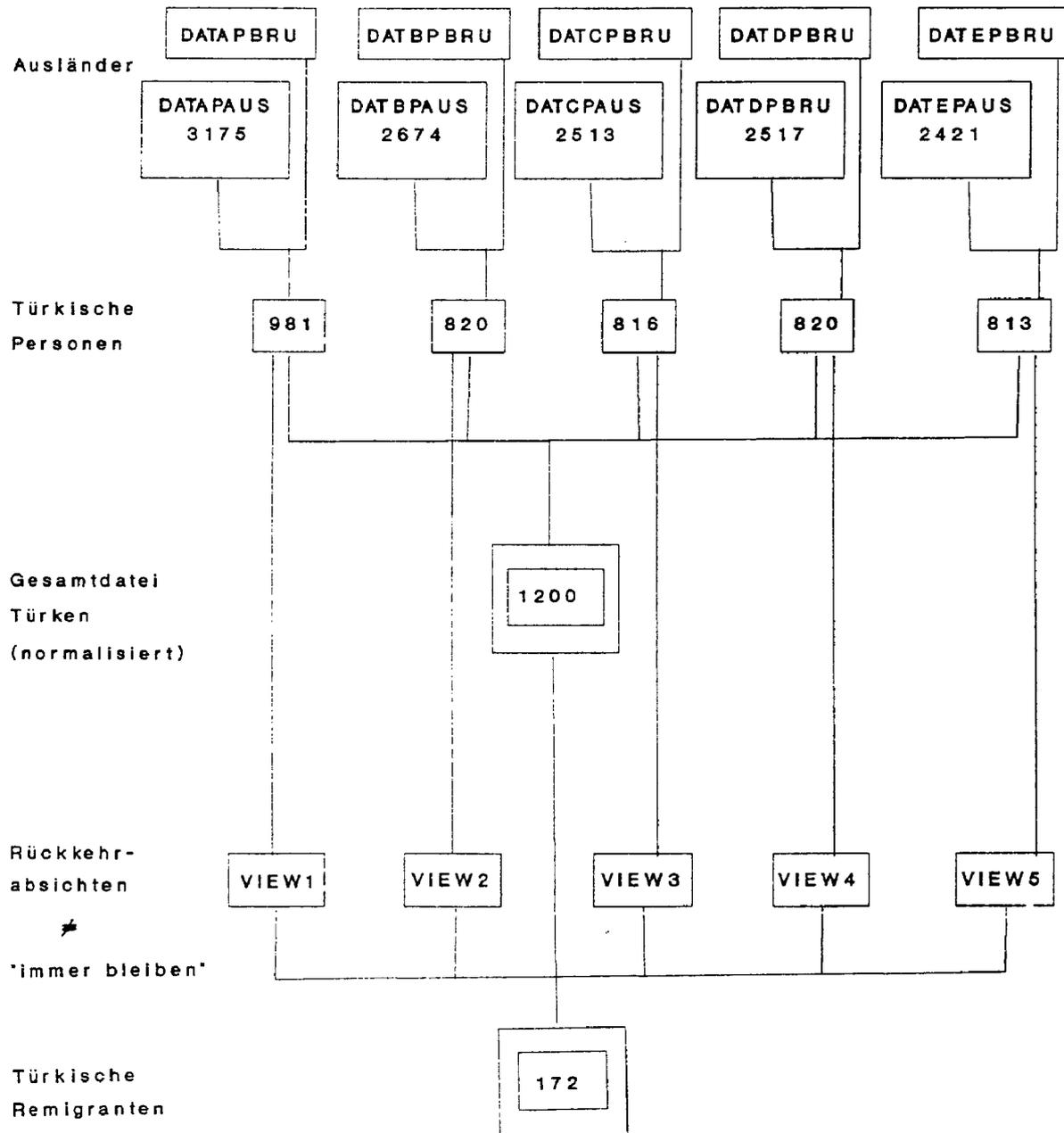
Um die Ergebnisse aus den Verknüpfungen weiterverwenden zu können, lassen sich VIEWS (d.h. Datensichten) oder neue Tabellen kreieren. Ein VIEW ist im Gegensatz zu einem TABLE keine physikalische Tabelle, sondern nur eine virtuelle, d.h. er erlaubt eine alternative Datensicht auf die Basistabellen, aber nicht die Möglichkeit, Daten einzutragen oder zu modifizieren. Ein VIEW dient in der Regel als Zwischenschritt für komplexe Abfragen; er wird aber auch oft aus Gründen des Datenschutzes verwendet, um einem bestimmten Benutzerkreis zwar Teilinformationen, aber nicht die Möglichkeit der Datenmanipulation zur Verfügung zu stellen. Bei großen Tabellen ist der Zugriff auf die verknüpfte Information über einen VIEW allerdings relativ langsam. Daher empfiehlt es sich, bei Verknüpfungen, die häufig benötigt werden, die Zwischenergebnisse nicht als VIEW, sondern in einer neuen Tabelle abzulegen, die jedoch je nach Datenmenge auch mehr Speicherplatz benötigt. Durch eine geeignete Speicherstruktur (s. Abschnitt 7) läßt sich aber die Abrufzeit der Daten nocheinmal verkürzen. Das Data Dictionary bietet einen Überblick, wie die jeweilige Objektstruktur gewählt wurde; z.B. kann man im Menüpunkt 'TABLES' das Untermenü 'EXAMINE' auswählen oder im SQL-Monitor über 'HELP objektname' die Definitionen erfahren.

Oft wird man auf die neu zusammengestellten Daten mit einem Statistikpaket o.ä. zugreifen wollen. Dazu muß man die Daten aus der INGRES-Datenbank auslagern, um sie auf der Arbeitsplatte am Großrechner zu halten. Am einfachsten läßt sich das Auslagern vornehmen, indem man im SQL-Monitor die entsprechende Datentabelle durch 'SELECT \* FROM tabelle;' auflistet und über den Menüpunkt 'FILE' die Datei gemäß der Konvention 'name.extention' abspeichert. Sie wird dadurch automatisch auf die Arbeitsplatte des Benutzers geschrieben, von wo sie in den XEDIT geladen und die überflüssigen Informationszeichen gelöscht werden können. Die so erstellte Datendatei kann sofort als Input für alle gängigen Statistikpakete (SAS, SPSSX, BMDP) verwendet werden.

Für die Analyse der Rückkehr von türkischen Gastarbeitern benötigt man vor allem die Informationen aus den Dateien DATxPAUSL (vercodete Fragen zum Personenfragebogen für Ausländer). Das Diagramm (Abbildung 3) zeigt die nötigen Verknüpfungen, die sowohl durch TABLES als auch durch VIEWS entstanden sind. Die so gewonnenen Daten aus dem SOEP wurden mit den

Statistikpaketen SPSSX und BMDP weiterverarbeitet.

Abbildung 3: Diagramm zur Extraktion von Information über türkische Gastarbeiter (Welle 1-5)



## 7 Datenbanktuning

INGRES erlaubt es jedem Entwickler, für die einzelnen Basistabellen die Speicherstruktur zu wählen, die für bestimmte Anwendungen optimal ist. Die Zugriffsmethode determiniert die benötigte Zeit des Datenbanksystems, um die gewünschten Daten zu finden und hat zudem Einfluß auf die Belegung von Plattenkapazität. Je umfangreicher eine Tabelle ist und je öfter sie verwendet wird, desto wichtiger wird der direkte schnelle Zugriff. Im wesentlichen gibt es vier Speicherstrukturen:

- heap: Für kleinere Tabellen ideal oder wenn alle Datensätze auf einmal gebraucht werden. Durch Löschen freigewordener Platz wird nicht wiederverwendet.
- hash: Vorteilhaft, wenn ein Primärschlüssel vollständig für den Zugriff auf genau einen Datenbanksatz angegeben werden kann.
- isam: Insbesondere geeignet, wenn die Tabelle nicht oder nur sehr langsam wächst und die Verteilung des Primärschlüssels fest bleibt.
- btree: Bei schnell oder stetig wachsenden Tabellen sowie bei häufiger Verwendung eines Datensatzes, der zu diesem Zweck sortiert vorliegen sollte, vorteilhaft.

Da auf die Basistabellen der Panel-Datenbank sehr oft zugegriffen wird, ist die Speicherstruktur 'btree' zu empfehlen. Mit der Anweisung

```
MODIFY tabellenname  
TO Speicherstruktur (7)  
ON spalte;
```

wird der Basistabelle eine bestimmte Speicherstruktur zugewiesen, wobei 'spalte' den Schlüssel (für die Panel-Datenbank der Primärschlüssel 'HHNR') angibt, nachdem die Datensätze sortiert abgelegt werden. Durch die Auflistung weiterer Spaltennamen kann der Schlüssel aus mehreren Spalten definiert werden, um den direkten Zugriff zu den Datensätzen möglichst eindeutig zu erhalten, so z.B. die Variable 'PERSNR' für die Panel-Datenbank, wenn mit personenbezogenen Tabellen gearbeitet wird.

Als weitere Verbesserung der Ablaufgeschwindigkeit, die allerdings den Maßnahmen wie sorgfältige Konzeption beim Datenmodellentwurf und richtige Wahl der Speicherstruktur und Zugriffsmethoden nachgeordnet sind, sind die Einführungen von Sekundärindizes über Indextabellen

zu nennen, die auf denselben Tabellenspalten verschiedene Zugriffsmethoden nebeneinander ermöglichen, sowie im Multi-User-Betrieb das Verwenden von Sperrmechanismen.

## 8 Die Datensicherheit bei INGRES

Nur eingetragene INGRES-Benutzer, die vom Datenbankadministrator für den Zugang zum INGRES-Datenbanksystem autorisiert sind, können ihre privaten Tabellen und Objekte einrichten und damit arbeiten. Wird auf eine Datenbank durch mehrere Benutzer zugegriffen, muß eine zusätzliche Autorisierung durch den Datenbankbesitzer eingerichtet werden. Die Objekte werden für diese Zugangsberechtigungen explizit angegeben, geeigneterweise können eingerichtete VIEWS, die nur die nötigen Informationen für die weiteren Benutzer enthalten, ausgewählt werden. Bei der Vergabe der Privilegien kann zwischen 'insert', 'update', 'delete', 'select' und 'all' gewählt werden, wobei 'select' den maximalen und 'all' den geringsten Zugriffsschutz gewährleistet.

Ein Hinweis zur Datenkontrolle und zur Datensicherung:

Gerade bei der Eingabe der Tabellenstruktur liegt eine große Fehlerquelle. Insbesondere die Formatangaben müssen exakt stimmen, damit beim Importieren der Daten den Variablen die richtigen Ausprägungen zugeordnet werden. Eine Kontrollmöglichkeit besteht darin, die Summe der Formatanzahl, die im Menüpunkt TABLES/EXAMINE als Reihenbreite mitausgewiesen wird, mit der angegebenen Satzlänge bei der Bandsatzbeschreibung für jede Datei zu vergleichen. Bei gesplitteten Dateien muß für die Summenbildung der mehrfache Variablen 'kopf' herausgerechnet werden.

Werden mit der Datenbank viele Anwendungen ausgeführt und neue Objekte erzeugt, sollte in regelmäßigen Abständen ein Backup der Datenbank angefertigt werden. Bei der Größe der Panel-Datenbank ist es erforderlich, das Backup auf ein separates Band auszuführen. Dazu legt man als ersten Zwischenschritt eine ausreichend große temporäre Platte an, auf die mittels der Befehlsfolge

```
UNLOADDB datenbankname (FMODE temp. Platte
UNLOAD
```

 (8)

die Dateien ausgelagert werden. Von dieser Platte können die Backup-Dateien im herkömmlichen Verfahren auf ein Band ausgelesen werden.

Sind Objekte der Datenbank gelöscht oder neu eingerichtet worden, sollte auf der Systemebene

der Befehl

*SYSMOD datenbankname* (9)

verwendet werden, um den freigewordenen Platz wieder verfügbar zu machen und die Datenbank zu reorganisieren. Die Ergänzung der Datenbank durch neue Panel-Wellen erfolgt völlig problemlos nach dem geschilderten Verfahren, solange genügend Speicherplatz zur Verfügung steht.

INGRES ist ein in die Zukunft gerichtetes Datenbanksystem. Sowohl das Arbeiten im Netzwerk als auch mit verteilten Datenbanken wird durch INGRES/NET bzw. INGRES/STAR ermöglicht. Fremde Datenbanksysteme lassen sich durch sogenannte Gateways integrieren, z.B. kann damit auf dBase-Applikationen zugegriffen werden. Die Portabilität erlaubt die Übertragung von Anwendungen auf eine Vielzahl von Hardware- und Softwareumgebungen; INGRES gibt es nicht nur für MS-DOS und VM/CMS, sondern auch für VAX-VMS, Ultrix, Unix, Hp, Sun, Apollo und weitere Systeme.

Das neue INGRES Release 6 verfügt darüber hinaus über eine Multi-Server-Architektur, die die optimale Ausnutzung von Mehrprozessorsystemen sowie das Einrichten mehrerer Datenbankserver pro CPU ermöglicht und somit eine wesentlich höhere Verarbeitungsgeschwindigkeit erreicht. Für die Zukunft ist geplant, das Ablegen von Bildern in der Datenbank – bedeutungsvoll etwa für den CAD/CAM-Bereich – zu erlauben sowie regelbasierende Anwendungen zu unterstützen.

## References

- [] Berger, Peter: Datenbankabfrage mit SQL. München, 1989
- [] Date, C.J.: A Guide to Ingres. Addison-Wesley-Publishing Comp., 1987
- [] Deutsches Institut für Wirtschaftsforschung (Hrsg.): Das Sozio-ökonomische Panel. Benutzerhandbuch, Version 3-Dezember 1989. Berlin, 1989
- [] Frick, Joachim: Die SIR-Datenbank des Sozio-ökonomischen Panels. - Ein Tutorial zu Aufbau, Syntax und problemorientierten Anwendungen. DIW, Version 90.1, Berlin, 1990
- [] Hanefeld, Ute: Das Sozio-ökonomische Panel. Grundlagen und Konzeption. Frankfurt/New York, 1987
- [] Infratest Sozialforschung: Das Sozio-ökonomische Panel, Welle 5. Methodenbericht zur Haupterhebung. München, 1989
- [] Lenhardt, Günter: Strukturierung und Auswertung von Mikrodatenfiles mit Hilfe von Datenbanksystemen. SFB3-Arbeitspapier Nr.56, Universität Frankfurt, 1981
- [] Parthier, Ulrich: Ingres - Datenbankmanagement der Zukunft. CHIP-Special, Würzburg, 1989
- [] Relational Technology Inc.: INGRES Reference Guide. Alameda, 1987
- [] van der Lans, Rick F.: Das SQL-Lehrbuch. Addison-Wesley Verlag, 1988