

METHODEN – VERFAHREN – ENTWICKLUNGEN

Nachrichten aus dem Statistischen Bundesamt

Ausgabe 1/2002

Methoden der Bundesstatistik – Weiterentwicklung

Komplexe Anwendungen mit Blaise for Windows: Erhebungsmanagement
und Datenbereinigung für die „Laufenden Wirtschaftsrechnungen“ 3

Blaise in der Zeitbudgeterhebung 2001/02
– Computergestützte Datenerfassung und Datenprüfung 14

Europa-Echo

Aktuelle Konjunkturdaten für Europa
– Zwei neue Serviceangebote des Eurostat Data Shop Berlin 20

Buchbesprechung

Peter von der Lippe: „Chain Indices – A Study in Price Index Theory“ 21

Herausgeber: Statistisches Bundesamt, Wiesbaden

Fachliche Informationen

zu dieser Veröffentlichung:

Gruppe IB,
Tel.: 06 11 / 75 20 77
Fax: 06 11 / 75 39 50
christian.koenig@destatis.de

Allgemeine Informationen

zum Datenangebot:

Informationsservice,
Tel.: 06 11 / 75 24 05
Fax: 06 11 / 75 33 30
info@destatis.de
www.destatis.de

**Veröffentlichungskalender
der Pressestelle:**

www.destatis.de/presse/deutsch/cal.htm

Erscheinungsfolge: (in der Regel) halbjährlich

Erschienen im März 2002

© Statistisches Bundesamt, Wiesbaden 2002

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.

Methoden der Bundesstatistik – Weiterentwicklung

Komplexe Anwendungen mit Blaise for Windows: Erhebungsmanagement und Datenbereinigung für die „Laufenden Wirtschaftsrechnungen“*)

1 Vorbemerkung

Blättert man durch die Dokumentationen, die im Anschluss an die Blaise Users Conferences entstanden sind, so findet man schnell drei recht umfassende Beiträge zum *Thema household expenditure survey*. Der Schwerpunkt der Betrachtung variiert dabei vom Einsatz von Codiertechniken (Roessingh/Bethlehem 1993), über den Vergleich der Erhebungsverfahren PAPI und CATI (Manners 1995) bis zum Einsatz von Blaise III als Software für die Integration verschiedener Erhebungen (Manners/Deacon 1997). Auch Schwamb und Wein (1997) haben sich zu diesem Thema geäußert. Angesichts dieser Veröffentlichungen stellt sich die Frage, ob es überhaupt, und wenn ja, was es zur *household expenditure survey* aus Sicht der Blaise-Programmierer und -Anwender noch Neues zu berichten gibt.

Der Begriff „neu“ kann hier unter verschiedenen Blickwinkeln betrachtet werden: Zum einen gibt es die technische Sicht, ob z.B. neue Möglichkeiten der Blaise-Entwicklungswerkzeuge genutzt werden. Hier auf werde ich nur cursorisch eingehen. Zum anderen kann die besondere Berücksichtigung der Anwender mit ihren speziellen Bedürfnissen und Aufgaben bei der Entwicklung einer Software zu einer neuen Sichtweise führen. Genau auf diesen Punkt ist dieser Aufsatz fixiert: Ist es möglich, mit Blaise die statistischen Fachabteilungen so zu unterstützen, dass auch komplexe Bedingungen und überaus heterogene Anforderungen abgedeckt werden? Um die Planungen und deren Realisierungen in Form eines *Erfahrungsberichtes* verständlich zu machen, sollen zunächst die statistischen Grundlagen beschrieben werden; sodann wird der Anforderungskatalog der Statistiker in seinen Grundzügen aufgeblättert und zu guter Letzt die Lösung präsentiert.

2 Statistische Grundlagen

Die Statistik der Laufenden Wirtschaftsrechnungen wird in Deutschland seit 1950 regelmäßig durchgeführt. Bundesweit nahmen zunächst 1 000 Haushalte, nach der Wiedervereinigung ab 1993 weitere 1 000 Haushalte auf freiwilliger Basis an der Erhebung teil, die auf drei Haushaltstypen fußte: Arbeitnehmerhaushalte mit mittlerem Einkommen sowie Beamten- und Angestelltenhaushalte mit höherem Einkommen zu jeweils vier Personen und Zwei-Personen-Haushalte von Renten- und Sozialhilfeempfängern. Die Haushalte führten ein monatliches Haushaltsbuch, in dem alle Einnahmen und Ausgaben minutiös festgehalten wurden. In der Regel verblieben die Haushalte mehrere Jahre in der Stichprobe, um eine möglichst kontinuierliche Datenlieferung sicherzustellen. Die Datenaufbereitung in den statistischen Ämtern erfolgte mit traditionellen Mitteln: Die eingegangenen Hefte wurden manuell geprüft, die Signierung der vom Haushalt erworbenen Produkte erfolgte handschriftlich in die Bücher. Nachdem die zentrale Datenerfassung die umfangreichen Erhebungsdaten in ein Großrechnerverfahren aufgenommen hatte, wurden die Daten in Batchläufen und manuellen Zwischenschritten bereinigt. Die Erhebungsverwaltung erfolgte mit Karteikarten.

Hatte sich diese Vorgehensweise in den 60er und 70er Jahren bewährt, so geriet sie spätestens in den 90er Jahren ins technisch-organisatorische Abseits: Aufgrund der nicht unbeträchtlichen Belastung der Haushalte sank deren Teilnahmebereitschaft, was wiederum eine abnehmende Repräsentativität zur Folge hatte. Die Zeitspanne von der Erhebung bis zur Veröffentlichung wurde als zu lang angesehen, der Aufbereitungsprozess galt als wenig effektiv und wenig effizient.

*) Artikel und Präsentation sind Herrn Bernward Hausmann gewidmet, der während seiner aktiven Dienstzeit im LDS NRW den Einsatz von Blaise stets gefördert hat.

Die deutsche Statistik beschloss daher 1994, mit einem neuen methodischen Konzept diese Schwierigkeiten zu überwinden. Folgende Umstrukturierungen sollten vorgenommen werden:

- Neue Erhebungsbücher sollten den Haushalten erleichtern, den umfangreichen Fragenkatalog zu beantworten.
- Um die Belastung der teilnehmenden Haushalte zu vermindern, wurde ein rollierendes Quartalspanel entworfen. Nunmehr sollten 6 000 Haushalte in der Stichprobe enthalten sein; jeder Haushalt brauchte aber nur einmal im Quartal anzuschreiben. Damit sollte der gesetzlich vorgegebene Stichprobenumfang (2 000 Haushalte) gewahrt bleiben und gleichzeitig eine breitere Streuung der Haushalte erreicht werden.
- Die Schichtung der Stichprobe sollte anhand anderer soziodemographischer Fakten, die aus dem Mikrozensus abgeleitet wurden, ebenfalls eine neue Basis und damit eine größere Repräsentativität erhalten.
- Die technische Unterstützung sollte auf ein völlig neues Fundament gestellt werden. Die Batchverfahren waren durch Dialoglösungen abzulösen, wobei eine umfangreiche Unterstützung aller Arbeitsprozesse mit DV-Mittel anzustreben war.

In einer Pilotstudie wurde 1996 und 1997 die Machbarkeit der Neukonzeption untersucht (vgl. Chlumsky/Ehling 1997, Gertkemper 1997). Aufgrund der positiven Ergebnisse erhielt das LDS NRW 1998 vom statistischen Verbund den Auftrag, die erforderlichen Programme für das neue Verfahren zu erstellen.

3 Anforderungen an das DV-System

Üblicherweise steht am Beginn eines DV-Projektes eine Auftragsanalyse, um die fachlichen und organisatorischen Anforderungen der späteren Anwender zu ermitteln. Als Ansprechpartner standen dem vierköpfigen Entwicklerteam zwei Fachleute zur Verfügung, die zum einen aus dem eigenen statistischen Amt, zum anderen aus dem Statistischen Bundesamt kamen. Da beide Ämter aufgrund von internen Umstrukturierungen im Bereich der Wirtschaftsrechnungen einen Personalwechsel vorgenommen hatten, konnte auf persönliches Know-how, welches im Rahmen der Testerhebung erworben worden war, nicht zurückgegriffen werden. Es stellte sich überdies schnell heraus, dass die für die Testerhebung entstandenen Programme aus verschiedenen Gründen weitgehend nicht weiter verwendet werden können. Kurzum: Es galt ein ganz neues DV-System zu entwerfen und zu programmieren. Glücklicherweise ergänzten sich die Vorstellungen der Fachvertreter in idealer Weise: Während das Statistische Bundesamt besonderen Schwerpunkt auf Stichprobenziehung, Plausibilitätsprüfungen und Ergebniserstellung legte, konzentrierte sich die Sichtweise der Vertreterin des LDS auf die Erhebungsorganisation und praktische Nutzbarkeit der Programme. Es kristallisierten sich alsbald neun Bausteine heraus, die zu erstellen waren:

1. Adressenverwaltung der interessierten und teilnehmenden Haushalte,
2. Stichprobenziehung,
3. Erhebungsmanagement,
4. Erfassung und Plausibilisierung der statistischen Angaben,
5. Aufbau und Pflege eines Klassifikationssystems (COICOP),
6. Budgetierung der Einnahmen und Ausgaben der Haushalte,
7. Typisierung der Haushalte,
8. Export der bereinigten Daten zur Ermittlung der Gewichtungsfaktoren,
9. Reimport der Gewichtungsfaktoren sowie Export für die Tabellierung.

Für die Testerhebung hatte das LDS NRW ein Basismodul für die *Adressenverwaltung* geschrieben. Dieses Modul konnte in der Grundstruktur zwar übernommen, musste aber um wesentliche Bestandteile erweitert werden. Da für die Erhebung auf die vorliegenden Anschriften der Haushalte zurückgegriffen werden sollte, die bereits zuvor den Berichtskreis gebildet oder an der Einkommens- und Verbrauchsstichprobe teilgenommen hatten und überdies mit Werbemaßnahmen weitere Haushalte gewonnen werden sollten, mussten sowohl ein Importprogramm sowie eine Dublettenprüfung „eingebaut“ werden. Um die Suche in den Adressbeständen zu vereinfachen, wurden neben der Haushaltsnummer als Primärschlüssel weitere Sekundärschlüssel generiert. Selbstverständlich wurden das Postleitzahl- sowie das Bankleitzahlverzeichnis als coding file sowie diverse Exportformate integriert, um die Weiterverarbeitung der Adressdaten in Form von Serienbriefen, Etikettenaufklebern usw. zu erleichtern.

Während sich diese Maßnahmen noch ohne weiteres mit den Standardlösungen von Blaise realisieren ließen, so wurde bei der Programmierung der *Stichprobenziehung* offenbar, dass das Prinzip der Datenhaltung mit Blaise ein Redesign des technischen Konzeptes erforderlich machte. Die den Laufenden Wirtschaftsrechnungen zugrunde liegende Stichprobe wird als Quotenstichprobe gebildet. Die Quoten beruhen auf Kombinationen der folgenden Merkmale:

Übersicht 1
Bei der Quotenbildung verwendete statistische Merkmale

Haushaltstyp	Haushaltsnettoeinkommen DM	Erwerbstätigkeit
Einpersonenhaushalt	unter 1 000	ja
Ehe/Lebenspartnerschaft ohne Kinder	1 000 bis unter 2 500	
Alleinerziehend	2 500 bis unter 5 000	nein
Ehe/Lebenspartnerschaft mit Kindern	5 000 bis unter 7 000	
Sonstiger Haushalt	7 000 und mehr	

Durch Auskombinieren der Merkmalsausprägungen ergeben sich maximal 50 Schichten, von denen allerdings 15 mit anderen Schichten zusammengefasst werden, so dass insgesamt 35 Schichten zu besetzen sind. Abhängig von der Bevölkerungszahl und -struktur der Länder legt der Quotenplan die Anzahl Haushalte je Schicht fest, wobei eine erhebliche Spannweite gegeben ist: Von den insgesamt 6 000 Haushalten entfallen z.B. auf NRW 1 050 Haushalte; die größte Schicht enthält 90 Haushalte, während andere mit weniger als fünf Haushalten belegt sind. Die Haushalte in jeder Schicht müssen innerhalb des Quartals gleichmäßig verteilt werden, um saisonale Klumpungseffekte zu vermeiden. Sofern der Quotenplan nicht vollständig erfüllt werden kann, dürfen nach einer detaillierten Abbildungsvorschrift Haushalte mit ähnlichen Schichtmerkmalen ersatzweise zugewiesen werden.

Es scheint naheliegend, hierfür mehrere Blaise-Files anzulegen, um mit einem einfachen Algorithmus den Vorgang zu steuern: Sobald ein Haushalt aufgrund seiner Strukturmerkmale einer Schicht zugeordnet und dies gesichert worden ist, wird in den anderen Files eine entsprechende Änderung vorgenommen, die aus Gründen der Konsistenz ebenfalls gespeichert werden muss. Genau an dieser Stelle kann das Dateisystem von Blaise keine Unterstützung bieten: Es ist nicht zulässig, gleichzeitig mehrere Dateien zum Lesen und Schreiben geöffnet zu halten, wie dies bei allen PC-Datenbanken ohne weiteres möglich ist. Da in der Zeit von Blaise III API-Schnittstellen noch Zukunftsvisionen darstellten, andererseits aber im Blaise-System gearbeitet werden sollte, mussten alternative Lösungen gefunden werden. Letztlich ließ sich der komplexe Vorgang der Stichprobenziehung in eine Abfolge sequenzieller Schritte umwandeln, in denen zunächst aus der Grundmasse die Haushalte zur Stichprobe zugewiesen werden, deren Merkmale mit der Schichtungsstruktur exakt übereinstimmen. Alle nicht gezogenen Haushalte werden sogenannte Reservehaushalte. In einem zweiten Schritt wird aus diesem Bestand

die Stichprobe ersatzweise aufgefüllt. Nach dieser Zuordnung werden die Haushalte gemäß Quotenplan auf die sogenannten Monatswellen aufgeteilt. Protokollierungen belegen diese Vorgänge sehr detailliert, so dass die Fachstatistiker sich jederzeit einen genauen Überblick über die Erfüllung der Quotenvorgaben verschaffen können.

Ein wesentlicher Bestandteil des *Erhebungsmanagements* ist die Verwaltung der teilnehmenden Haushalte. Die Haushalte wirken freiwillig an dieser umfangreichen Statistik mit. Eine sorgsame Pflege des gewonnenen Bestandes ist ebenso bedeutsam für die Kontinuität des Berichtskreises wie schnelles und flexibles Reagieren bei Rückfragen, Problemen und Änderungswünschen. Für das Erhebungsmanagement erwachsen daraus mehrere Aufgaben:

- Die Adressen müssen aktuell sein.
- Es gilt, den Versand der Unterlagen an die teilnehmenden Haushalte vorzubereiten und auszulösen.
- Wollen Haushalte in einem Quartal pausieren, in einem anderen Monat das Haushaltsbuch führen oder aus der Erhebung ausscheiden, muss dies unmittelbar registriert werden können und zu entsprechenden Konsequenzen beim Versand usw. führen.
- Hat ein Haushalt seine Unterlagen abgeliefert, ist ein entsprechender Eingangsvermerk zu setzen, damit unnötige Nachfragen vermieden werden.
- Andererseits soll es möglich sein, ausstehende Unterlagen anzumahnen.
- Die teilnehmenden Haushalte erhalten eine bescheidene Prämie. Die Auszahlung dieser Beträge muss unmittelbar möglich sein und über einen beleglosen Zahlungsverkehr abgewickelt werden können.
- Es muss jederzeit ein aktueller Zustandsbericht über den Stand der Aufbereitung erstellt werden können. Diverse Geschäftsstatistiken sind dafür zu implementieren.

Rückgrat für all diese Aktivitäten ist die sogenannte Haushaltsdatei (siehe Abbildung 1) als zentrales Steuerelement. Darüber hinaus wird sie auch für weitere wichtige Arbeitsprozesse innerhalb der Laufenden Wirtschaftsrechnungen automatisch im Hintergrund herangezogen. Die hierfür erforderlichen Felder des Datenmodells bleiben dem Anwender im Dialog verborgen.

Wie sehr vom Erhebungsmanagement Flexibilität verlangt wird, kann am Beispiel der Prämienzahlung verdeutlicht werden. Da jedes statistische Amt bei der Auszahlung andere Rahmenbedingungen setzen will, nämlich die Abhängigkeit der Prämienauszahlung von Dateneingang und Bearbeitungsstand, musste eine dynamisch konfigurierbare Prüfbedingung programmiert werden:

Übersicht 2
Für die Prämienberechnung dynamisch zu setzende Parameter

Eingang der Unterlagen	Logische Verknüpfung	Bearbeitungsstand
Ja	Und	In Bearbeitung Nachfrage Editing abgeschlossen
Nein	Oder	Budgetierung abgeschlossen Jahresrechnung abgeschlossen Versand an Statistisches Bundesamt erfolgt

Die entsprechenden Parameter werden in einer Blaise-Datei (INIT) gesetzt, im Rahmen eines Manipulationslaufes eingelesen und beeinflussen die Auszahlung der Prämien.

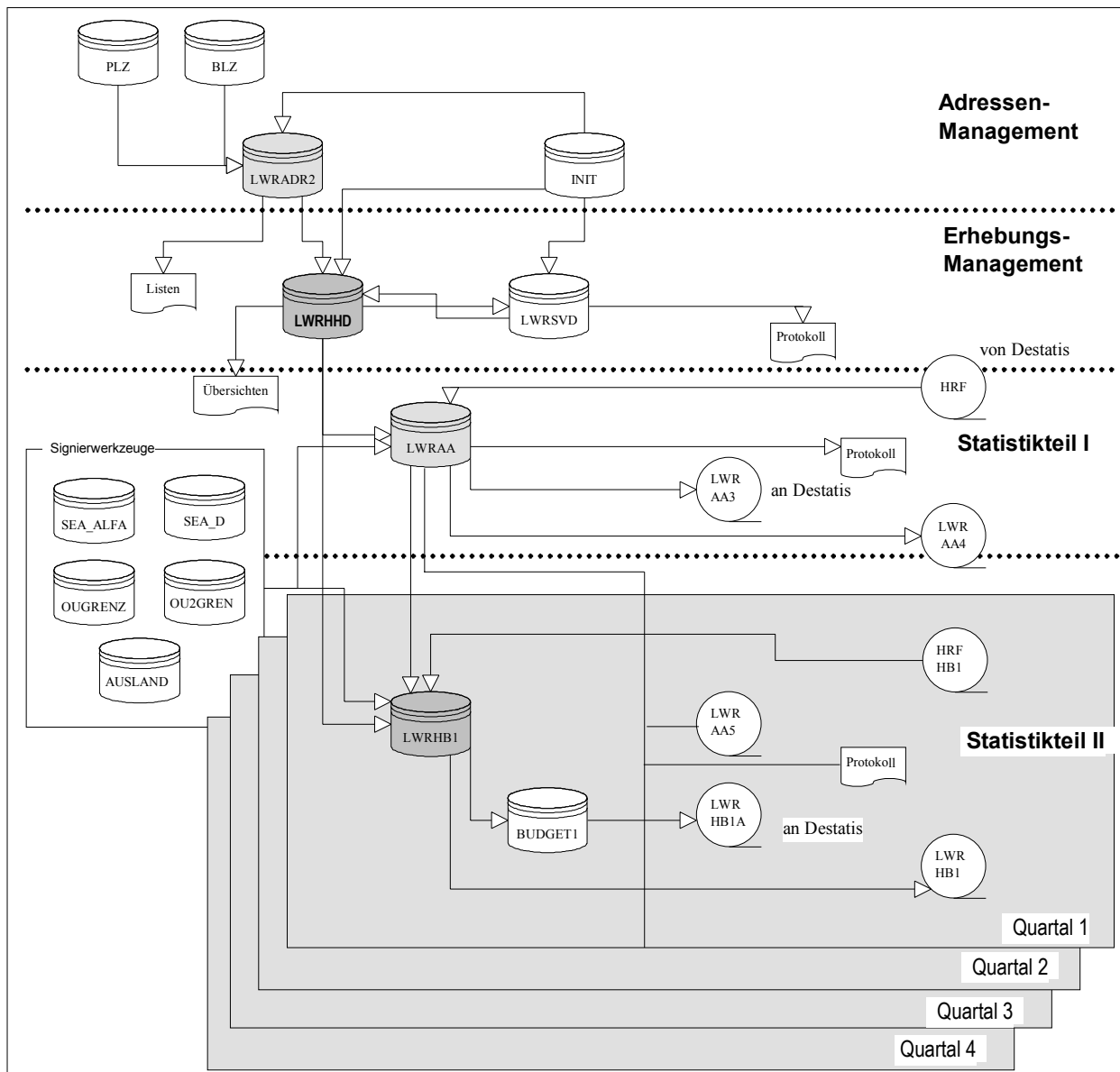
Das Erhebungsmanagement betrifft nicht nur die Zusammenarbeit mit den Haushalten. Es müssen auch innerhalb der statistischen Ämter die Arbeitsprozesse gesteuert werden. Dies spiegelt sich wider in einer Mitarbeiterverwaltung, welche die unterschiedlichen Bearbeitungsrechte der Anwender regelt: Während der Supervisor Zugriff auf alle Funktionen hat und damit der eigentliche „Herr des Verfahrens“ ist, kann der Mitarbeiter nur spezielle Arbeitsschritte, z.B. Eingabe der Fragebogen nebst Fehlerbereinigung, ausführen. Die statistischen Ämter setzen für die Datenbereinigung auch sogenannte Heimarbeiter ein, die ausgestattet mit technischem Equipment der Ämter zu Hause die Haushaltsbücher erfassen und bearbeiten. Zu diesem Zweck werden mit Hilfe des Erhebungsmanagements Arbeitsportionen gebildet; dies beinhaltet auch die aus Gründen des Datenschutzes erforderliche faktische Anonymisierung der Adressinformationen.

Fachlicher Kernpunkt der Anwendung sind natürlich die Module zur *Erfassung und Plausibilisierung* der statistischen Angaben. Die Laufenden Wirtschaftsrechnungen bestehen aus zwei Teilen. Im ersten Teil, den sogenannten *Allgemeinen Angaben*, werden Informationen über die soziale Stellung der Haushaltsmitglieder, deren Erwerbstätigkeit, zur Haushaltsgröße und zum Nettoeinkommen, zur Wohnsituation und über die Ausstattung mit Gebrauchsgütern erhoben. Das zwölfseitige Heft wird den Haushalten zum Jahreswechsel zugeschickt und bis Ende Januar zurückerwartet.

Aus technischer Sicht stellt dieser Fragebogen keine besonderen Anforderungen. Da die Liste der vorhandenen Gebrauchsgüter jährlich modifiziert wird, wurde hier mit Hilfe eines *external file* eine flexible Zugriffslogik geschaffen, welche aus der Gesamtmenge der in Frage kommenden Güter die im jeweiligen Erhebungsjahr zutreffenden auswählt und anzeigt.

Den zweiten Erhebungsteil bilden die sogenannten *Haushaltsbücher*, in denen die Haushalte im Detail sämtliche Einnahmen und Ausgaben sowie alle Veränderungen, die im Haushalt seit dem letzten Berichtszeitraum eingetreten sind, nachweisen. Auf die statistisch-fachliche Komplexität dieses Erhebungsteils soll nur mit einigen Zahlen verwiesen werden: Die Fachseite hat 200 Fehlerspezifikationen erstellt, in die praktisch alle vorkommenden Felder involviert sind. Da bis zu zehn Personen je Haushalt erfasst werden können und Abhängigkeiten zwischen den einzelnen Haushaltsmitgliedern bestehen (Eltern – Kinder, Einkommensbezieher – unterstützte Person usw.), lassen sich die logischen Erfordernisse und der damit verbundene Programmieraufwand leicht ermessen.

Abbildung 1
Schematische Übersicht über die Datenmodelle



Zunächst einmal stellt die Ablaufsteuerung des Erhebungsprozesses eine Herausforderung dar:

- Aus dem Erhebungsteil *Allgemeine Angaben* müssen grundlegende soziodemographische Merkmale so fortgeschrieben werden, dass sie für die Plausibilisierung der Haushaltsbücher herangezogen werden können. Dieser Arbeitsgang ist allerdings von der erreichten Datenqualität abhängig zu steuern, überdies muss er beliebig oft wiederholt werden können und alle Quartale berücksichtigen, ohne die Datenintegrität zu gefährden.
- Um die Daten auswerten zu können, müssen bundesweit Hochrechnungsfaktoren ermittelt werden. Zu diesem Zweck sieht das Fachkonzept vor, dass der plausible Datenbestand eines jeden Quartals von allen Landesämtern zentral bereitgestellt wird. Das Statistische Bundesamt ermittelt die Hochrechnungsfaktoren für die gesamte Stichprobe und schickt sie fallbezogen den Ländern zurück. Diese wiederum importieren die Faktoren in den Datensatz und können nunmehr Auswertungen vornehmen. Es ist bei dieser Vorgehensweise sicherzustellen, dass niemand mehr die Daten

nach dem Export verändern kann; jede Modifikation, z.B. im sozialen Status, hätte Auswirkungen auf die Hochrechnung. Es wurde daher ein Konzept entworfen und realisiert, das beim Export der Daten diese gewissermaßen „versiegelt“. Derart gekennzeichnete Datensätze können nur noch betrachtet, aber nicht mehr verändert werden. Während mit Blaise III keine systemseitigen Lösungen für diese Anforderung verfügbar waren und umständlich gesonderte Views und Browser programmiert werden mussten, konnte ab der Windows-Version (4.3) über den READ ONLY-Parameter endlich eine zeitgemäße Zugriffslogik realisiert werden.

- Das „Versiegelungsverfahren“, welches einen größtmöglichen Schutz der Daten vor irrtümlicher Veränderung bietet, stellt sich inzwischen in Einzelfällen als Hindernis bei der Fehlerbeseitigung heraus. Erst im Rahmen der Auswertungen wurden Fehler entdeckt, die auf unzureichenden Spezifikationen beruhen. Dank Versiegelung war der Zugriff auf die Daten allerdings gesperrt. Um bei derartigen Problemfällen dennoch Abhilfe zu schaffen, musste ein spezielles Verfahren zum Aufheben der Versiegelung eingeführt werden. Aus Sicherheitsgründen steht es nur dem Systemverwalter zur Verfügung.

Für das Entwicklerteam wurde mit der Programmierung der Haushaltsbücher eine neue Dimension eröffnet: Jeder Datensatz umfasst mehr als 7 200 Felder und belegt 35 KByte Plattenplatz. Dieses Volumen beruht neben den umfangreichen personenbezogenen Angaben zu den Einnahmen insbesondere auf dem Teil der Haushaltsbücher, in denen die täglichen Ausgaben verzeichnet werden. Den Haushalten stehen hier auf 19 Seiten über 600 Zeilen zur Niederschrift zur Verfügung, wobei neben der genauen Artikelbezeichnung und dem Betrag festgehalten wird, ob die Ausgabe bar oder unbar, in DM oder Euro, im In- oder Ausland getätigt wurde. Es lag nahe, diese Fragen als Tabelle umzusetzen, in welche die leistungsstarken coding tools von Blaise integriert werden sollten. Bei der Realisierung und im Testbetrieb traten allerdings knifflige Probleme auf:

- Da innerhalb der Ausgaben spezielle Prüfungen vorgenommen werden sollten – hat z.B. ein Mieter Ausgaben für Miete angegeben –, sollte dies in einer einzigen, auf 600 Zeilen angelegten Tabelle geschehen. Die Fachseite wünschte außerdem zu jeder Ausgabe eine Kannfehlerprüfung auf Einhaltung von Unter- und Obergrenzen. Obwohl 1999 schon recht leistungsstarke PCs zur Verfügung standen, dauerte der Aufbau der Tabelle auf dem Bildschirm erheblich zu lange. Wartezeiten von einer Minute waren nicht selten. Zur Beschleunigung wurde zunächst die Wertebereichsprüfung so umgestaltet, dass sie nur bei ausdrücklicher Anforderung angewendet wird. Inzwischen wurde sie ganz entfernt. Die Erhebungspraxis zeigt überdies, dass normalerweise kaum mehr als 300 bis 400 Eintragungen erforderlich sind. Der Anwender kann daher vor dem Aufruf der Tabelle in einer Variablen definieren, wie viele Zeilen er einzugeben beabsichtigt. Außerdem kann er mit einer Abbruchbedingung (Ende oder Code 9999999) die Tabelle vorzeitig verlassen (vgl. Abbildung 2).

Abbildung 2
Eingabe „999999“ zum Abbruch der Dateneingabe

	Code	nb	HBAusTag	eu	HBAudtag	Ausl
HB_27[1]	0540351	<input type="checkbox"/>	334,00	<input type="checkbox"/>	334,00	<input type="checkbox"/>
HB_27[2]	011	<input type="checkbox"/>	33,00	<input type="checkbox"/>	33,00	<input type="checkbox"/>
HB_27[3]	0312361	<input type="checkbox"/>	67,00	<input type="checkbox"/>	67,00	<input type="checkbox"/>
HB_27[4]	0942350	<input checked="" type="checkbox"/>	437,00	e	854,70	<input type="checkbox"/>
HB_27[5]	999999	<input type="checkbox"/>				<input type="checkbox"/>
HB_27[6]		<input type="checkbox"/>				<input type="checkbox"/>
HB_27[7]		<input type="checkbox"/>				<input type="checkbox"/>

- Die Codierung der täglichen Ausgaben beruht auf SEA-Klassifikation (Systematisches Verzeichnis der Einnahmen und Ausgaben der privaten Haushalte), die aus der COICOP/HBS-Systematik (Classification of Individual Consumption by Purpose/Household Budget Surveys) abgeleitet ist. Es war beabsichtigt, diese Klassifikation dem Bearbeiter sowohl als hierarchischen Suchbaum als auch als alphabetisch sortierte Liste nebst trigram coding anzubieten. Bei der Umwandlung der SEA-Klassifikation in einen Blaise coding file stellte sich allerdings heraus, dass der hierarchische Suchmechanismus unter Blaise III fehlerhaft arbeitet und nicht verwendet werden konnte. Der SEA-Schlüssel enthält als erste Hierarchiestufe einen zweistelligen Schlüssel (01, 02, ... 11, 12). Aufgrund der Sortierverfahren des coding tools wird die Reihenfolge der Schlüssel vertauscht (z.B. 02 nach dem Schlüssel 12), da die führende Null offenbar nicht berücksichtigt wird.

Um mit der alphabetischen Suche eine möglichst große Trefferquote zu gewährleisten, wurden die Texte der SEA-Klassifikation analysiert, zerlegt und neu aufbereitet, so dass sie nun mit einem Kurztext und einem erläuternden Langtext bei der Recherche angeboten werden (vgl. Abbildung 3). Dieses Vorgehen hat sich sehr bewährt; bei der Eingabe der Warenbezeichnung werden auf Anhieb ca. 90% der Schlüssel unmittelbar gefunden. Die Trefferquote erhöht sich somit um fünf Prozent beim Wechsel in den Trigrammodus.¹⁾

Ursprünglich sollten die coding files jederzeit ergänzt werden können, um z.B. neue Gebrauchsgüter oder auch Synonyme aufzunehmen. Da im dafür erforderlichen Read-write-Modus im Netzbetrieb mit mehreren Nutzern ganz erhebliche Performanzprobleme auftraten, wurde mit Erfolg auf den Read-only-Modus umgestellt. Überdies können die Anwender die coding files auch auf die lokalen Maschinen übertragen und dort direkt zugreifen, wenn sich das LAN als zu langsam erweisen sollte. Inzwischen werden alle Neuerungen und Ergänzungen für die Klassifikation zentral gesammelt, fachlich und technisch aufbereitet und als unmittelbar einsetzbare coding files in den statistischen Verbund verschickt. Das LDS NRW hat dem Statistischen Bundesamt für diesen Zweck eine eigene Anwendung zur Pflege der SEA-Register übergeben.

1) Es wäre sehr praktisch, wenn unter B4W die Suchbegriffe beim Wechseln des Suchmodus im Eingabefeld erhalten blieben, wie es unter Blaise III noch der Fall war!

Abbildung 3
Alphabetische Signierung über den Kurztext

Kurztext	Orig_code	Klartext
CD-Brenner	0913015	CD-Brenner, PC-Zubehör
CD-Hüllen	0914030	CD-Hüllen
CD-Player	0911141	CD-Player (ohne solche für Kraftfahrzeuge)
CD-Player Kfz	0911143	CD-Player für Kraftfahrzeuge (Radio-Kassetten- oder Radio-CD-Player-Kombinationen für Kraftfahrzeuge - 0911131)
CD-ROMs	0914013	CD-ROMs, unbespielt, für PC
CD-ROMs	0914024	CD-ROMs, bespielt (ohne System- und Anwendungssoftware)
CD-ROMs	0931012	CD-ROMs, bespielt mit Videospiele
CD-Ständer	0511059	CD-Ständer
CDs	0914013	CDs, unbespielt
CDs	0914021	CDs, bespielt
Cembali	0922130	Cembali
Cervelatwurst	011	Cervelatwurst, Rohwurst
Champagner	021	Champagner (ohne Kantinen- und Gaststättenverzehr)
Champagner	1111057	Verzehr von Champagner in Restaurants, Cafés, Bars, an Imbissständen
Champagner	1112050	Verzehr von Champagner in Kantinen, Mensen

Suchen: Schlüsseltypen alfa Seacode tri

Auf die Verfahren zur *Budgetierung*, zur *Typisierung* der Haushalte sowie zum *Export* der bereinigten Daten sowie zum *Reimport* der Gewichtungsfaktoren soll hier unter fachlichen Gesichtspunkten nicht weiter eingegangen werden. Die für diese Aufbereitungsschritte erforderliche technische Unterstützung wurde in umfangreichen Manipula-Programmen realisiert. Obschon in NRW nur 1 050 Haushalte in jedem Quartal zu bearbeiten sind, benötigen diese Programme aufgrund ihrer fachlichen Komplexität und der extensiven Datenmodelle z.T. erhebliche Laufzeiten. Glücklicherweise hat der Umstieg auf B4W beim Datenexport unerwartete Performanzsteigerungen mit sich gebracht: Liefen diese Programme unter Blaise III etliche Stunden, so sind sie nun in wenigen Minuten abgearbeitet. Für die erforderlichen abschließenden Plausibilitätsläufe (C/S, vgl. unten) trifft dies aber leider nicht zu, mit der Konsequenz, dass während dieser Zeit ein Dialogbetrieb nicht möglich ist.

4 Generelle Aspekte

Die Komplexität der gesamten Anwendung lässt sich sicherlich daran messen, dass über 20 z.T. sehr umfangreiche datamodels definiert und knapp 60 manipula setups generiert werden mussten. Es ist dem Entwicklerteam gelungen, alle Arbeitsprozesse so zu modellieren und programmtechnisch zu implementieren, dass sie innerhalb der Anwendung gestartet werden können. Der Wechsel von Blaise III nach B4W ermöglichte überdies die Integration des MS Office-Paketes. Auswertungen und graphische Darstellungen der Geschäftsstatistiken können ohne weiteres von den Anwendern in eigener Regie angestoßen werden. Allerdings war der Aufwand für die Entwicklung des Konzeptes, die Umsetzung, die Tests und die Einführung groß. Insgesamt wurden mehr als sechs Entwicklerjahre²⁾ investiert. Die Akzeptanz bei den Anwendern in allen statistischen Ämtern ist gut; im Zuge einer Langzeitbeobachtung wurde ermittelt, dass alle Anforderungen praxisrelevant umgesetzt worden sind. Trotz dieser Anerkennung sind aus technisch-organisatorischer Sicht kritische Anmerkungen zu machen.

Das DEP-Modul von Blaise bietet grundsätzlich die Möglichkeit, zwischen verschiedenen Modi der Dateneingabe sowie der entsprechenden Darstellung am Bildschirm zu wählen. Da seinerzeit die Entwicklung mit Blaise III begonnen wurde, konnten zur Fehlerkennzeichnung im Edit-Modus nur die recht unauffälligen error counter verwendet werden. Im reinen Interview-Modus können Mussfehler aber nicht unterdrückt werden, um sie nach Rückfragen später zu korrigieren. Um dennoch eine zügige Dateneingabe im Interviewmodus mit voller Fehlermeldung und der Option, diese zu unterdrücken, zu

2) Mein Dank gilt B. Wlodkowski, F. Merks, A. Hansen, M. Broose und dem Blaise-Support vom CBS NL.

ermöglichen, musste nach anderen Wegen gesucht werden. In der bereits erwähnten INIT-Datei wird eine Variable mit der Prüfvorgabe *check/signal* (C/S) vorbelegt. Das gesamte Routing der datamodels wird so aufgebaut, dass alle Prüfungen in einen Programmablauf eingebunden sind, der abhängig von der Prüfvorgabe entweder checks oder signals auslöst (vgl. Übersicht 3).

Übersicht 3 Code-Beispiel für die extern gesteuerte dynamische Prüfung

```
{Check or signal is activ until new command}

IF exini.hb_prufe[q] = muss THEN      {muss: var. in INIT.DBD; muss = hard error}
    CHECK
ELSE
    SIGNAL
ENDIF

IF (nb = EMPTY) AND (eu <> EMPTY) THEN
    exini.jahr >= 2002
INVOLVING(eu, nb)
    "No EURO-cash before 2002!"
ENDIF

IF ((Code = '2099201') OR (Code = '2099301')) AND (nb = '5') THEN
    eu = EMPTY
    "[S19B*] Bei Eingabe der gemieteten Garagen- und Stellplatzanzahl darf die
    Spalte 'Euro' nicht mit 'x' oder 'e' signiert werden."
ENDIF
```

Die Anwender unterdrücken im Interesse raschen Arbeitens zumeist alle Fehler. Da aber sichergestellt werden muss, dass nur fehlerfreies Material in die weitere Verarbeitung kommt, muss ein abschließender Prüflauf erfolgen, in dem alle Muss-Fehler ermittelt werden. Dieser Vorgang macht sich ebenfalls die beschriebene Lösung zu Nutze: Automatisch wird in der INIT-Datei auf check umgestellt, im Batch werden alle Sätze geprüft. Gleichzeitig wird eine einfache Fehlerstatistik erstellt. Bedauerlicherweise blockieren diese Läufe das System nachhaltig – eine Verweildauer von drei bis vier Stunden ist nicht selten, so dass sie meistens nachts gestartet werden.

Diese Performanzprobleme in einigen Prozessen der Anwendung werden von den Anwendern moniert. In NRW nutzen normalerweise etwa zwölf Mitarbeiter die Anwendung. In Spitzenzeiten wird allerdings der Mitarbeiterstab verdoppelt, so dass sich alle Störungen nachhaltig auf den Aufbereitungsprozess auswirken. Hier wären Verbesserungen beim Abarbeiten von Plausibilitätsprüfungen in großen Tabellen sowie beim handling of transactions im Blaise-Laufzeitsystem sehr willkommen.

Ein bislang ungelöstes Problem stellen gelegentliche Datenverluste dar: Beim Abspeichern eines Datensatzes stürzt die Anwendung auf der lokalen Arbeitsstation ab, die Eingaben zu einem Haushalt, die mitunter zwei bis drei Stunden beansprucht hatten, sind partiell, manchmal aber auch vollständig verloren. Unseligerweise sind keine Regelmäßigkeiten zu erkennen, so dass die Fehleranalyse schwierig ist. Da die eingesetzten Programme offenkundig syntaktisch korrekt sind und im Prinzip auch funktionieren, scheint das Blaise-System als Fehlerverursacher auszuschneiden. Das LDS hat sich intensiv bemüht, mögliche andere Ursachen für diese Abstürze herauszufinden. Zunächst stand die physische Struktur des Netzes auf dem Prüfstand. Nach dem Umzug in ein neues Dienstgebäude mit einer optimalen technischen Ausstattung traten die Probleme aber weiterhin auf. Auch die Hinweise des CBS, Netzwerkparameter anzupassen, haben keine grundlegende Verbesserung bewirkt. Zur Zeit wird mittels einer genauen Anschreibung für jeden eingesetzten PC der Versuch unternommen, ggf. lokale Rechnerspezifika als störauslösende Elemente einzugrenzen.

Eine letzte Bemerkung bezieht sich auf den Versionswechsel von Blaise III nach B4W. Die Ankündigung, dass per Knopfdruck die DOS-Programme nach Windows konvertiert werden würden, löste im Vorfeld bei den Entwicklern Skepsis aus. Selbst das Blaise-Team war a priori nicht überzeugt, dass der Umstieg problemlos zu bewältigen sei. Zwar verflog diese Skepsis nach den ersten erfolgreichen Versuchen, insgesamt musste aber doch mehr Arbeit investiert werden, als zuvor angenommen worden war. Zum einen hing dies mit Verbesserungen zusammen, die sich aus den neuen Möglichkeiten ergaben (Anschluss an MS Office), zum anderen mussten alle Module neu gewandelt, geprüft und in die Applikation eingebunden werden. Bei dieser Fleißarbeit waren Schwierigkeiten mit der Standard-Konvertierungsroutine (B4W Version 4.2/4.3) für den Wechsel von ASCII nach ANSI ebenso hinderlich wie die stellenweise lückenhafte Dokumentation. Dank des Supports des CBS konnten die Umstellungsprobleme aber letztlich gut gelöst werden.

5 Schlussfolgerung

Um die eingangs gestellte Frage, ob es möglich sei, mit Blaise komplexe Anwendungen im Dienste der Statistik zu erstellen, zu beantworten: Natürlich lautet die Antwort „ja“. Es hat sich gezeigt, dass insbesondere nach der Umstellung auf die Windows-Version die heterogenen fachlichen Anforderungen praktisch vollständig befriedigt werden konnten. Die Blaise-Entwicklungsumgebung enthält zahlreiche leistungsstarke Werkzeuge, um einerseits eine enge Verzahnung von DV-Unterstützung und statistischem Aufbereitungsprozess zu ermöglichen, andererseits kann aber ebenso ein Erhebungsmanagement mit vollständiger Adressen-, Haushalte-, Mitarbeiter- und Stichprobenverwaltung realisiert werden, das keine Anwenderwünsche offen lässt. Es lohnt sich insofern sehr, den skizzierten integrativen Ansatz auch für andere Statistiken weiterzuverfolgen. Allerdings wäre es dann wünschenswert, bestimmte Module – z.B. die Adressen- und Mitarbeiterverwaltung – als Basisbausteine in generalisierter, jederzeit wiederverwendbarer Form bereitzustellen. Das Telefonmanagementsystem von Blaise ist ein ausgezeichnetes Beispiel für ein derartiges Vorgehen. In der gemeinsamen Arbeit des statistischen Verbundes in Deutschland zeichnen sich erste Entwicklungen in diese Richtung ab; wünschenswert wäre es, wenn derartige Module für alle Blaise-Anwender zur Verfügung stünden, ein lohnenswerter Auftrag für das CBS NL.

Literaturhinweise

Chlumsky, J.; Ehling, M. (1997): Grundzüge des künftigen Konzepts der Wirtschaftsrechnungen der privaten Haushalte, in: *Wirtschaft und Statistik* 7, S. 455 – 461.

Gertkemper, F. (1997): Ergebnisse der Testerhebung zum neuen Konzept der Laufenden Wirtschaftsrechnungen, in: *Wirtschaft und Statistik* 12, S. 872 – 878.

Manners, T. (1995): The UK Family Expenditure Survey: comparison of PAPI and CAPI estimates from a controlled trial using Blaise, in: *Essay on Blaise 1993. Proceedings of the Third International Blaise Users Conference*, pp. 101 – 114.

Manners, T.; Deacon, K. (1997): An Integrated Household Survey in the UK: The Role of Blaise III in an Experimental Development by the Office For National Statistics, in: *Actes de la 4^e Conférence Internationale des Utilisateurs de Blaise 1997*, pp. 197 – 211.

Roessingh, M.; Bethlehem, J. (1993): Trigram coding in the Family Expenditure Survey of the CBS, in: *Essay on Blaise 1993. Proceedings of the Second International Blaise Users Conference*, pp. 118 – 132.

Schwamb, H.J.; Wein, E. (1997): Notes on an Integrated Data Management, in: *Newsletter 10*, ed. by International Blaise User Group, London, pp. 33 – 35.

*Thomas Pricking, Landesamt für Datenverarbeitung und Statistik Nordrhein-Westfalen (LDS NRW),
Tel.: 02 11 / 94 49 50 58, E-mail: pricking@lds.nrw.de*

Blaise in der Zeitbudgeterhebung 2001/02 – Computergestützte Datenerfassung und Datenprüfung

1 Hintergrund

Datenbedarf

Die Ergebnisse der Zeitbudgeterhebung werden vielfach genutzt. Dies gilt bis heute für die erste Erhebung von 1991/92 und wird auch für die zweite, gegenwärtige Erhebung 2001/02 gelten:

- in Projekten der Bundesministerien als Grundlage für politische Entscheidungen,
- von Wissenschaftlerinnen und Wissenschaftlern im In- und Ausland, die mit den anonymisierten Zeitverwendungsdaten arbeiten,¹⁾
- ad hoc für spezielle Anfragen etwa von Zeitschriften-, Rundfunkredaktionen oder interessierten Bürgerinnen und Bürgern,
- in Veröffentlichungen des Statistischen Bundesamtes, die ein breites Spektrum von Themen behandeln.²⁾

Methoden der Datenerhebung

Diese Nutzungsmöglichkeiten setzen eine gute Datenbasis voraus. Bei der Gewinnung einer Datenbasis können unterschiedliche Formen der Computerunterstützung zum Einsatz kommen:

- Die Befragten geben in ihrem Haushalt einem anwesenden Interviewer Auskunft, der die Angaben unmittelbar in einen tragbaren Computer eingibt (Computer Assisted Personal Interview – CAPI).
- Die Befragten geben per Telefon einem Interviewer Auskunft (Computer Assisted Telephone Interview – CATI).
- Die Befragten füllen selbst die Fragebogen zu Hause am Computer aus (Computer Assisted Self Interviewing – CASI).
- Die Befragten beantworten einen im Internet bereitgestellten Fragebogen (Computer Assisted Web Interviewing – CAWI).
- Die Erfasserinnen und Erfasser in den Statistischen Ämtern des Bundes und der Länder geben nachträglich Daten in Blaise ein, die traditionell mit Papierfragebögen erhoben wurden (Computer Assisted Data Input – CADI).

Die aktuelle Zeitbudgeterhebung 2001/02 ist als Selbstausfüllererhebung konzipiert. Ob anstelle von gedruckten Erhebungsunterlagen, die bei der ersten Erhebung 1991/92 zum Einsatz kamen, bei der neuen Erhebung auch eine internetbasierte Zeitanschreibung realisierbar wäre, sollte ein im Juni 2000 gemeinsam mit der Universität Lüneburg veranstalteter Workshop mit Wissenschaftlerinnen und Wissenschaftler aus Forschungsinstituten und aus der amtlichen Statistik klären. Ernüchternd war ein zentrales Ergebnis dieses Workshops: Ein direktes Ausfüllen der Erhebungsunterlagen via Internet durch die Befragten selbst ist für derart große Zeitbudgeterhebungen zur Zeit noch nicht verlässlich genug, um die Repräsentativität der Daten und somit allgemeingültige Aussagen zur aktuellen Zeitverwendung in Deutschland zu gewährleisten. Insbesondere bei einem Vergleich mit der Erhebung von

1) Für einen Einblick in das Forschungsspektrum siehe Ehling, M., Merz, J. u.a. (2001): Zeitbudget in Deutschland – Erfahrungsberichte der Wissenschaft, Bd. 17 der Schriftenreihe Spektrum Bundesstatistik, hrsg. vom Statistischen Bundesamt, Wiesbaden.

2) Siehe z.B. Blanke, K., Ehling, M., Schwarz, N. (1996): Zeit im Blickfeld. Ergebnisse einer repräsentativen Zeitbudgeterhebung, Band 121, Schriftenreihe des Bundesministeriums für Familie, Senioren, Frauen und Jugend, Stuttgart, Berlin, Köln oder Holz, E. (2000): Zeitverwendung in Deutschland – Beruf, Familie, Freizeit –, Bd. 13 der Schriftenreihe Spektrum Bundesstatistik, 1. unveränd. Nachdruck 2001, hrsg. vom Statistischen Bundesamt, Wiesbaden.

1991/92 wäre dann nicht mehr abzuschätzen, ob Unterschiede in der Zeitverwendung auf das neue Erhebungsinstrument oder auf tatsächliche gesellschaftliche Veränderungen während der letzten zehn Jahre zurückzuführen wären (die Vorträge des Workshops sind als Veröffentlichung erhältlich, siehe Fußnote 1).

Datenerfassung mit Blaise

Daher werden für die Erhebung 2001/02 gedruckte Unterlagen an die Haushalte verschickt und nachträglich computerunterstützt erfasst (CADI-Verfahren). Geplant ist ein Stichprobenumfang von 5 000 Haushalten³⁾. Die Statistischen Landesämter werben und betreuen die teilnehmenden Haushalte, sichten die eingegangenen Erhebungsunterlagen und fragen in Zweifelsfällen bei den Haushalten nach. Die Datenerfassung erfolgt für alle Bundesländer zentral beim Statistischen Bundesamt mit Hilfe des Programmpakets Blaise für Windows, das vom Statistischen Amt der Niederlande entwickelt wurde. Die Erhebungsunterlagen bestehen aus einem Haushaltsfragebogen, einem Personenfragebogen und einem Tagebuch. Entsprechend wurden drei Blaise-Programme entwickelt, die wesentliche Vorteile einer computergestützten Datenerfassung wie automatische Filterführung und Plausibilitätskontrollen ausnutzen.

Plausibilitätskontrollen mit Blaise

„Plausibilitätskontrolle“ („PL“) bedeutet vereinfacht gesagt, dass die Eingaben dahingehend überprüft werden, ob sie im realen Leben vorkommen können oder ob es sich um Erhebungsfehler handelt. Manche Angaben sind undenkbar und müssen daher auf jeden Fall korrigiert werden (daher auch der Name „Muss-Fehler“), andere Eintragungen sind zwar denkbar, wenn auch nicht sehr wahrscheinlich (sog. „Kann-Fehler“). Blaise gewährleistet als CADI-Verfahren eine begleitende, permanente Plausibilitätskontrolle während der Datenerfassung. Dies ermöglicht eine schnelle Reaktion auf etwaige Unklarheiten.

2 Das Blaise-Programm für den Haushaltsfragebogen

Aufbau des Haushaltsprogramms

Das Haushaltsprogramm dient zur Erfassung von Angaben über die Zusammensetzung des Haushalts, über die Wohnsituation, über die erhaltenen Hilfeleistungen und über das Haushaltseinkommen. Darüber hinaus gewährleistet der Registrierschlüssel und das Hilfsmerkmal „Vorname“ die Verknüpfung von Haushaltsinformationen einerseits sowie Personen- und Tagebuchinformationen andererseits. Der in Abbildung 1 dargestellte Programmauszug veranschaulicht die Funktionsweise am Beispiel der Fragen zur Haushaltszusammensetzung.

3) Zu Methode und Organisation siehe Ehling, M., Holz, E., Kahle, I. (2001): Erhebungsdesign der Zeitbudgeterhebung 2001/2002, in: *Wirtschaft und Statistik*, 6, S. 427 – 436.

Beispiel für einen Kann-Fehler im Fragenkomplex zum Haushaltseinkommen: Es erscheint wenig wahrscheinlich, dass Kindergeld die Haupteinkommensquelle eines Rentnerhaushalts ist. Unwahrscheinlich ist jedoch nicht gleichzusetzen mit unmöglich, so dass im Einzelfall überprüft werden muss, ob eine Korrektur der Angabe nötig ist.

Neben der automatischen PL ist auch die automatische Filterführung ein Vorteil des Haushaltsprogramms. So springt Blaise nach der Erfassung der Beziehungsangaben direkt zum nächsten inhaltlichen Fragenkomplex, der Wohnsituation. Die beschriebene Anwendung kann ohne Programmierkenntnisse bedient werden. Die weitgehende Anlehnung an die gedruckte Vorlage erleichtert die Orientierung innerhalb des Programms.

3 Erfassung der Zeitverwendung – Das Blaise-Programm für die Tagebücher

Die Erfassung der Tagebücher, die von den Befragten völlig frei geführt werden, ist neben dem Haushaltsfragebogen ein wesentliches Kernstück der Programmierung in Blaise. Die Tagebücher werden durch ein Blaise-Programm in Form von „Zeitabschnitten“ bzw. „Zeitblöcken“ erfasst. Ein „Zeitabschnitt“ bzw. „Zeitblock“ ist eine Zeitspanne mit gleichbleibenden Eintragungen der Befragten zur Hauptaktivität, dem Verkehrsmittel bzw. dem Ort, der gleichzeitigen Aktivität und mit wem die Zeit verbracht wurde. Jeder Zeitabschnitt hat gemäß den 10-Minuten-Intervallen des Tagebuchs eine Uhrzeit, zu der er begonnen und eine, zu der er beendet wurde. Ein vorhergehender Zeitabschnitt ist beendet und ein neuer beginnt, falls sich bei einer oder mehreren Eintragungen etwas ändert. Ein neuer Zeitabschnitt mit eigener Start- und Endzeit beginnt also auch, wenn die Hauptaktivität gleich bleibt, aber Änderungen bei den anderen Tagebucheintragungen zu erkennen sind. Eine Änderung ist auch gegeben, wenn bisherige Eintragungen nicht mehr vorliegen, etwa eine gleichzeitige Aktivität wegfällt oder bei den Mit-wem-Merkmalen kein Kästchen mehr angekreuzt ist.

Ein Tagebuchtag: Aufbau und Umsetzung in Blaise

Alle Teilnehmerinnen und Teilnehmer an der Zeitbudgeterhebung 2001/02, die mindestens 10 Jahre alt sind, füllen an drei Tagen ihre Tagebücher aus, an zwei Tagen unter der Woche und an einem Tag am Wochenende. Die Tage müssen nicht aufeinanderfolgen. Ein „Tagebuchtag“ entspricht hierbei nicht einem Kalendertag von 0.00 bis 24.00 Uhr, sondern er beginnt um 4.00 Uhr in der Frühe und endet um 4.00 Uhr am Morgen des nachfolgenden Tages. Diese Vorgehensweise – im Einklang mit entsprechenden Eurostat-Empfehlungen zu Zeitbudgeterhebungen – verbessert die Auswertungsmöglichkeiten von Aktivitäten, die nach Mitternacht fortgeführt werden (z. B. von Arbeitszeiten oder Freizeitaktivitäten während der Nacht). Während dieser 24 Stunden werden die Befragten gebeten, im 10-Minuten-Rhythmus die von ihnen ausgeübte Hauptaktivität und gleichzeitige Aktivität in eigenen Worten einzutragen, ob und welches Verkehrsmittel sie benutzen sowie, falls zutreffend, eines oder mehrere Kästchen anzukreuzen, mit wem zusammen sie die Zeit verbringen (mit „Kindern unter 10 Jahren“, „dem/der (Ehe-)Partner/-in“, „anderen Haushaltsmitgliedern“, „sonstigen bekannten Personen“).

Hat beispielsweise eine Teilnehmerin der Zeitbudgeterhebung in ihr Tagebuch geschrieben, dass sie zwischen 16.00 und 16.50 Uhr ununterbrochen ein Buch las, so bedeutet dies für die Erfassung in Blaise, dass der Zeitabschnitt bzw. Zeitblock um 16.00 Uhr beginnt und um 16.50 Uhr endet, für diese Zeitspanne wird dann die Tätigkeit „Bücher lesen“ mit dem Aktivitätscode „813“ kodiert. Abbildung 2 zeigt den entsprechenden Teil der Eingabemaske in Blaise für das Tagebuch.

Abbildung 2
Blaise-Eingabemaske für das Tagebuch (Auszug)

Uhrzeit_Anfan	16:00	Uhrzeit_End	16:50
Hauptaktiv_eh	813	Text_Hauptak	
Ort_Verkehr	11	zu_Hause	
gleichz_Aktiv		Text_gleichz_	
mit_wem_ea[1]	2	Nein	
Kinder_emk[12]			

Um die Dateneingabe mit Blaise weiter zu vereinfachen und Fehlerquellen auszuschalten, wurde bei der Programmierung berücksichtigt, dass die Endzeit eines Zeitabschnittes automatisch die Startzeit des nachfolgenden ist. Dadurch müssen lediglich die Uhrzeiten, wann die jeweiligen Zeitabschnitte beendet wurden, eingegeben werden. Des weiteren wird in der Abbildung deutlich, dass bei dem Erfassungsfeld „Ort_Verkehr“, in dem der Aufenthaltsort oder das benutzte Verkehrsmittel eingegeben wird, „11“ für „zu Hause“ steht. Das Feld „gleichz_Aktiv“ ist leer, da die Teilnehmerin keiner gleichzeitigen Aktivität nachgeht. Bei „mit_wem_ea“ wurde eingegeben, dass sie diesen Zeitabschnitt nicht mit anderen verbringt – weder mit Kindern, dem (Ehe-)Partner oder anderen Haushaltsmitgliedern, ebensowenig mit anderen ihr bekannten Personen –, was gleichsam als Filter im Programm wirkt, die einzelnen „Mit-wem“-Variablen nicht abzufragen.

Noch einige weitere Beispiele für Zeitabschnitte bzw. Zeitblöcke aus dem fiktiven Tagebuch der Teilnehmerin: Der nächste Zeitabschnitt dauert von 16.50 bis 17.30 Uhr; sie sieht in dieser Zeit fern (die Hauptaktivität hat sich also geändert). Zwar macht sie das nach 17.30 Uhr weiterhin, aber ihr Ehemann ist hinzugekommen – die entsprechenden Mit-wem-Kästchen wurden im Tagebuch angekreuzt –, wodurch eine Änderung zur vorherigen Situation stattgefunden und somit ein neuer Zeitabschnitt begonnen hat. Der nachfolgende Zeitabschnitt wird durch die nächste Änderung um 18.00 Uhr eingeleitet, da der Ehemann wieder gegangen ist.

Programmbegleitende Plausibilitätsprüfungen („PL“) für die Tagebücher

Durch die programmbegleitenden Plausibilitätsprüfungen („PL“) bei Blaise wird die Erfassung der Tagebücher unterstützt. Prüfungen im Tagebuch für Muss-Fehler – wie weiter oben dargestellt, ist es bei diesen erforderlich, die eingegebene Angabe zu korrigieren – sind z. B., dass nur bestimmte Ziffern als Aktivitätscodes zulässig sind, dass nicht der identische Code für Haupt- und gleichzeitige Aktivität gewählt werden kann, dass während dem Schlafen keine gleichzeitigen Tätigkeiten ausgeübt werden können, und falls Schlafen als gleichzeitige Tätigkeit geschieht, bestimmte Verkehrsmittel ebenfalls ausscheiden (man kann zwar im Bus oder als Beifahrer im Auto schlafen, nicht aber, wenn man selbst am Steuer sitzt). Die Angaben zu den Uhrzeiten, wann Aktivitäten angefangen bzw. beendet werden, müssen dem 10-Minuten-Rhythmus entsprechen – 16.03 Uhr kann nicht eingegeben werden, jedoch 16.00 oder 16.10 Uhr. Die Uhrzeit, wann ein Zeitabschnitt bzw. Zeitblock beendet wurde, kann nicht

vor jener liegen, zu der er begonnen wurde. Außerdem bedürfen Wegezeiten auch der Angabe eines Verkehrsmittels.

Die Plausibilitätskontrollen in der Zeitbudgeterhebung sind keine Frage der Weltanschauung. Bei der Behandlung von Kann-Fehlern ist also Vorsicht geboten und behutsam zu verfahren. Oberster Grundsatz ist: Die Befragten haben mit ihren Einträgen Recht, es sind ihre Angaben. Keinesfalls dürfen „Korrekturen“ nur deshalb bei der Dateneingabe vorgenommen werden, weil die Angaben eines Befragten dem eigenen Weltbild oder eigenen Erfahrungen widersprechen oder untypisch sind. Daher sind diese Prüfungen eher sparsam verwendet worden.

Die Eingabe der Tagebücher in Blaise: Mehr als Datenerfassung

Bei der Erfassung der Zeitbudgeterhebung mit Blaise geht es also nicht nur um die Eingabe der relativ standardisierten Angaben des Haushalts- und Personenfragebogens, sondern es ergibt sich eine besondere Situation durch die offenen Eintragungen aus den Tagebüchern, für die passende Aktivitätskategorien (aus einer Zusammenstellung von mehr als 200) zu finden sind. Hierfür kann es hilfreich sein, das gesamte Tagebuch und die „Ergänzenden Informationen zum Tagebuch“ (z. B. ob der Tagebuchtag wegen Krankheit oder Urlaub ungewöhnlich war) sowie den Haushaltsfragebogen und die Personenbogen im Auge zu behalten. Es handelt sich also um die Erstellung einer Datenbasis, weniger um Datenerfassung.

Erleichterte Erfassung durch Zeitabschnitte/-blöcke

Die Programmierung der Tagebucheintragung in Form von Zeitabschnitten bzw. Zeitblöcken ist bedeutend besser als die alternative Möglichkeit, die Eintragungen aus den Tagebüchern für jeden 10-Minuten-Zeittakt im Computer zu erfassen. Ein Tagebuchtag besteht aus 144 Zeittakten, so dass 1 008 Angaben für einen einzigen Tagebuchtag – macht 3 024 Angaben für alle drei Tage – eingegeben werden müssten. Dies ist besonders im Falle von Aktivitäten, die ohne Unterbrechung über längere Zeitspannen ausgeübt werden, lästig und mit unnötiger Arbeit verbunden. Für das bekannte Beispiel, in welchem die Teilnehmerin zwischen 16.00 und 16.50 Uhr ein Buch liest, wären bei der obigen Vorgehensweise dennoch für alle 10-Minuten-Takte in diesem Zeitraum die Angaben wiederholt, d.h. fünfmal einzugeben. Bei der eingesetzten Erfassung in Form von Zeitabschnitten bzw. -blöcken musste die Eingabe hingegen nur einmal gemacht werden. Auch für die späteren statistischen Auswertungen sind die Zeitabschnitte interessant; natürlich können diese dann auch in Zeittakte umgewandelt werden.

4 Bereitstellung der mit Blaise erzeugten Datenbasis

Nach der jeweils getrennten Erfassung durch Haushaltsprogramm, Personenprogramm und Tagebuchprogramm liegen drei Datendateien vor. Dies entspricht schon gleich zu Beginn dem Datenaufbau, wie er auch später für den neuen „Scientific Use File“ und „Public Use File“ verwendet werden wird. Eine spätere zeitintensive Umorganisation der Daten wird überflüssig. Denn auch hier ist Schnelligkeit gefragt: Der Scientific Use File wird wissenschaftlichen Einrichtungen in Deutschland für deren eigene Analysen zur Verfügung gestellt werden, der Public Use File ist jedem Interessierten zugänglich, im Unterschied zum Scientific Use File auch Forscherinnen und Forschern im Ausland. Aus Gründen des Datenschutzes werden beide Datenfiles adäquat anonymisiert werden. Hierdurch wird verhindert, dass Angaben einzelner Befragter erkannt werden können.

Die mit Blaise erzeugten Datensätze im ASCII-Format können mit verschiedenen Programmen (z.B. Excel, ACCESS, SPSS, SAS, Statspez) analysiert und dargestellt werden.

Thomas Baumann / Erlend Holz, Statistisches Bundesamt, Wiesbaden, Tel.: 06 11 / 75 37 80 bzw. 38 97, E-mail: thomas.baumann@destatis.de bzw. erlend.holz@destatis.de

Europa-Echo

Aktuelle Konjunkturdaten für Europa – Zwei neue Serviceangebote des Eurostat Data Shop Berlin

Mit dem „Prime-UP“-Service und der Online-Datenbank der „Euro-Indicators“ hat das Statistische Amt der Europäischen Gemeinschaften (Eurostat) ein neues Tor geöffnet, Interessenten mit aktuellen Konjunkturdaten zu versorgen.

„Prime-UP¹²“ wird in Form eines Jahresabonnements (Mindestpreis: 150 Euro) angeboten und beinhaltet die Lieferung der vom Interessenten ausgewählten Indikatoren. Die Datenlieferung mit einer garantierten Mindestleistung von zwölf Aktualisierungen pro Jahr erfolgt zeitgleich zur Pressemitteilung um 12:00 MEZ per E-Mail. Auf der Homepage Eurostats werden die Indikatoren in 19 Tabellen aufgezeigt, aus denen eine erste Auswahl getroffen werden kann. Über den Eurostat Data Shop Berlin (EDS Berlin) können die Kunden sich dann anschließend einen unverbindlichen Kostenvoranschlag erstellen lassen. Dazu ist es notwendig, beim Absenden des Anfrageformulars unter den Data Shops „Germany“ als Anbieter zu wählen. Die Datenlieferung mit den Tabellenbezeichnungen erfolgt in englischer Sprache.

Mit der Homepage „Euro-Indicators“ bietet Eurostat den Nutzern erstmals einen Online-Zugriff auf eine Datenbank mit Konjunkturanalysedaten der EU, der Eurozone und der einzelnen EU-Mitgliedstaaten an. Der Zugriff zu dieser Datenbank wird als Jahresabonnement zum Preis von 1 000 Euro pro Jahr (Einzelnutzer) angeboten und kann beim EDS Berlin beantragt werden. Die täglich aktualisierten Daten auf der englischsprachigen Homepage umfassen folgende Bereiche:

- Zahlungsbilanz,
- Konjunkturerhebungen,
- Verbraucherpreise,
- Außenhandel,
- Industrie, Handel und Dienstleistungen,
- Arbeitsmarkt,
- Geld- und Finanzindikatoren sowie
- Volkswirtschaftliche Gesamtrechnungen.

Auf diesen Internetseiten stehen Interessenten zusätzlich noch eine Reihe weiterer Informationen, u.a. mehr als 90 kostenlose Indikatoren, Metadaten sowie Pressemitteilungen, zur Verfügung.

Beide Serviceangebote können als Hyperlink über die nutzerfreundliche Homepage des EDS Berlin (www.eu-datashop.de) erreicht werden, die seit August 2001 einen Zugang zu den Produkt- und Serviceangeboten Eurostats bietet.

Der *EDS Berlin* ist ein statistischer Informations- und Beratungsdienst Eurostats in Zusammenarbeit mit dem Statistischen Bundesamt (Destatis). Er richtet sich an alle Nutzer in Deutschland, die harmonisierte statistische Daten über die Europäische Union (EU-15), die Eurozone, die 15 EU-Mitgliedstaaten sowie deren Partner- und Beitrittsländer benötigen. Die von Eurostat harmonisierten Daten sind in Datenbanken (über 180 Millionen Daten) zusammengeführt, aus denen sich die Kunden maßgeschneiderte Extraktionen anfertigen lassen können. Die Klassifizierungspläne mit einer detaillierten Gliederung der Datenbanken sowie die entsprechenden Anfrageformulare für die Erstellung eines unverbindlichen Kostenvoranschlags sind auf der Homepage des EDS Berlin abrufbar.

Neben den Datenbankauszügen bietet der EDS Berlin die Veröffentlichungen Eurostats in elektronischer Form (CD-ROM, PDF-Datei) oder als gedruckte Version an. Auf der EDS Berlin-Homepage erhält der Nutzer eine Kurzbeschreibung interessanter und aktueller Publikationen mit direkter Bestelloption.

Ausführliche Informationen zum EDS Berlin unter:

Statistisches Bundesamt

i-Punkt Berlin/ Eurostat Data Shop

Otto-Braun-Straße 70/ 72, 10178 Berlin,

Telefon: 018 88 / 644 94 27, Telefax: 018 88 / 644 94 30,

E-Mail: datashop@destatis.de, Internet: <http://www.eu-datashop.de>

Servicezeiten:

Montag bis Donnerstag: 8.00 – 17.00 Uhr, Freitag: 8.00 – 15.00 Uhr

Öffnungszeiten der Präsenzbibliothek:

Montag bis Donnerstag: 11.00 – 17.00 Uhr

Mareike Daum, Statistisches Bundesamt, Berlin, Tel.: 018 88 / 644 94 39,

E-mail: mareike.daum@destatis.de

Buchbesprechung

Peter von der Lippe: „Chain Indices – A Study in Price Index Theory“

In der Reihe „Spektrum Bundesstatistik“ ist das Buch „Chain Indices – A Study in Price Index Theory“ von Peter von der Lippe im Frühjahr 2001 erschienen (Peter von der Lippe: Chain Indices – A Study in Price Index Theory, hrsg. vom Statistischen Bundesamt, Band 16 der Schriftenreihe Spektrum Bundesstatistik, Verlag Metzler-Poeschel, Stuttgart, 2001, 291 Seiten, EUR 19,84, ISBN 3-8246-0638-0). Es befasst sich mit der Wahl des am besten geeigneten Indextyps, einem überaus wichtigen und aktuellen Problem der amtlichen Preisstatistik. Infolge der Empfehlung des Systems of National Accounts, 1993 (SNA, 1993), und des darauf aufbauenden Europäischen Systems der Volkswirtschaftlichen Gesamtrechnungen, Ausgabe 1995 (ESVG, 1995), in dem die Verwendung von Kettenindizes empfohlen wird, hat sich eine nicht unerhebliche Anzahl der europäischen nationalen statistischen Ämter dazu entschlossen, von der Berechnung direkter Indizes („Festbasisindizes“) auf Kettenindizes überzugehen.

Peter von der Lippe unterzieht die in der Literatur üblicherweise genannten Vorteile von Kettenindizes einer grundlegenden Überprüfung. Sehr ausführlich und wissenschaftlich äußerst fundiert setzt er sich mit den Argumenten der Befürworter von Kettenindizes auseinander und stellt diesen die Nachteile gegenüber.

Der Beitrag von Peter von der Lippe stellt eine bemerkenswerte Gegenposition zum „Mainstream“ der überwiegend angelsächsischen Literatur dar. Mit seiner englischen Publikation hat er eine solide Grundlage für die weitere wissenschaftliche Diskussion über den „richtigen“ Indextyp für die Inflationsmessung – auch auf internationaler Ebene – geschaffen.

Das Buch besteht neben einer Einführung aus folgenden acht Kapiteln:

- Towards a clear notion of a chain index,
- Elements of (direct) index theory,

- Analysis of properties of chain indices,
- Elements of a theory of chain indices,
- Deflation with chain indices,
- Arguments in favor of the chain index approach,
- Divisia index and its relation to chain index numbers,
- Official statistics and concluding remarks.

Ute Egner, Statistisches Bundesamt, Wiesbaden, Tel.: 06 11 / 75 41 55, E-Mail: ute.egner@destatis.de

In eigener Sache

Regelmäßiger Bezug von „Methoden – Verfahren – Entwicklungen“ per E-Mail

Nach Einstellung der Druckfassung im vergangenen Jahr geben wir „Methoden – Verfahren – Entwicklungen“ ausschließlich als Internetpublikation heraus. Um unsere bisherigen, aber auch neue Interessenten auf das Erscheinen einer neuen Ausgabe aufmerksam zu machen und ihnen einen schnellen und problemlosen Zugang zu ermöglichen, bieten wir eine Benachrichtigung per E-Mail und eine Aufnahme in unseren E-Mail-Verteiler für „Methoden – Verfahren – Entwicklungen“ an. Mittels Hyperlink kann bei Interesse sofort auf die entsprechende PDF-Datei zugegriffen werden.

Sollten Sie daran Interesse haben, bitten wir um eine kurze Mitteilung an folgende E-Mail-Adresse: dirk.schreiber@destatis.de.