

## **Datenerschließung in der Statistik**

Eberhard Mann (BA, IIIb)



## Datenerschließung in der Statistik

Leider wird in der Öffentlichkeit noch immer kolportiert, dass die Bundesanstalt für Arbeit auf einen Datenschatz sitzt, den sie interessierten Wissenschaftlern böswillig vorenthält. Das war nie der Fall. Die Bundesanstalt für Arbeit war für Wünsche von Wissenschaftlern immer aufgeschlossen.

Da sind zunächst einmal die regelmäßigen Standardauswertungen, die in den „Amtlichen Nachrichten der Bundesanstalt für Arbeit“, der Fachreihe „Arbeitsmarkt in Zahlen“ und im Internet unter der Adresse <http://www.arbeitsamt.de/hst/services/statistik/index.html> veröffentlicht werden. Die Statistik der Bundesanstalt für Arbeit macht auf Anfrage darüber hinaus entgeltpflichtige Sonderauswertungen. Im letzten Jahr waren es über 200. Zu rund einem Viertel wurden sie von Universitäten, Hochschulen und Instituten angefordert.

Zur Beschäftigung ist mittlerweile eine anonymisierte Mikrodaten-Stichprobe erzeugt worden, und zwar bisher für fünf Quartale, beginnend mit dem 31.3.1998. Bis Ende dieses Jahres sollen weitere sieben Quartale folgen. Externe Wissenschaftler (DIW) testen zurzeit die anonymisierten Datensätze. Ziel ist die laufende Bereitstellung anonymisierter Mikrodatensätze zur Beschäftigung am aktuellen Rand. Weitere Statistiken sollen einbezogen werden.

Erstmals Mitte 1999 hat die Statistik der Bundesanstalt für Arbeit Verbleibsquoten für die Teilnehmer an Maßnahmen der Arbeitsförderung vorgelegt und damit ein Evaluationsfieber ausgelöst. Die Verbleibsquoten geben darüber Auskunft, ob eine Absolventin oder ein Absolvent ein halbes Jahr später arbeitslos gemeldet war oder nicht. Mittlerweile gibt es diese Informationen für die Jahre 1998, 1999 und 2000. Mitte nächsten Jahres werden zumindest für die Großaggregate berufliche Fortbildung und Arbeitsbeschaffungsmaßnahmen Eingliederungsquoten folgen. Dann wissen wir, wie viel der Absolventen ein halbes Jahr später sozialversicherungspflichtig beschäftigt waren.

Um derartige Aussagen treffen zu können, müssen Daten unterschiedlicher EDV-Fachverfahren miteinander verknüpft werden. Das ist nicht einfach. Die Bundesanstalt für Arbeit musste Neuland betreten. Wir lernen dabei laufend dazu. Heute kann man sagen, dass wir auf einem guten Weg sind.

Daten aus unterschiedlichen Fachverfahren sind nicht von vornherein kompatibel. Nicht selten fehlt auch ein gemeinsamer Identifikator. An einigen Beispielen wird das deutlich:

Im Fachverfahren coSach, das für die Abwicklung der Maßnahmen verwendet wird, sind keine Teilnehmer erfasst. Die Teilnehmer werden über die Maßnahmennummer im Fachverfahren coArb ermittelt. Die Verknüpfungsmöglichkeit kann nicht bei Eingabe überprüft werden, da die Daten nicht nur in unterschiedlichen Datenbanken, sondern auch auf unterschiedlichen Rechnern gespeichert sind. Hinzu kommt, dass für die Maßnahme häufig ein anderes Arbeitsamt als für die Betreuung des Arbeitslosen zuständig ist.

Nimmt ein Arbeitsloser an einer Folgemaßnahme der beruflichen Weiterbildung teil, wird in der coArb die neue Maßnahmennummer eingetragen und damit die Verknüpfung zur ersten Maßnahme überschrieben. Dieser Mangel wird wohl erst im Data Warehouse zufriedenstellend behoben werden können.

Bei Arbeitsbeschaffungsmaßnahmen wird zwar die Verknüpfung über einen Teilnehmerdatensatz im coSach hergestellt. Aber auch hier kann die Richtigkeit der Verknüpfung – wie bei der beruflichen Weiterbildung – nicht bei Eingabe überprüft werden.

Das Fachverfahren coLei ist ausschließlich auf die Abwicklung der Zahlungen ausgerichtet. Es kennt nicht den Leistungsempfänger **für** die Zeit von bis, sondern lediglich den Empfänger von Leistungen **in** der Zeit vom bis. Gründe hierfür sind: verspätete Bewilligungen, Nachzahlungen, Abwicklung als Einmalzahlung.

Damit Daten über Leistungsempfang und Arbeitslosigkeit zeitraumbezogen verknüpft werden können, müssen Daten an verschiedenen Stellen des coLei-Verfahrens gesammelt und historisch geführt werden.

Hierzu muss man wissen, dass beim Gestalten neuer Geschäftsprozesse in der Vergangenheit die fachlichen Vorgaben im Wesentlichen vom zuständigen Fachbereich - etwa die Vermittlung oder die Leistung - und damit aus dessen Perspektive getroffen und von der IT-Organisation umgesetzt worden sind. Belange der Statistik, des übergreifenden Controlling, der Kostenrechnung, der Evaluation usw. sind gar nicht oder nicht ausreichend berücksichtigt worden. Belastbare Daten für die genannten Zwecke mussten und müssen – soweit das überhaupt möglich ist - mühsam nachträglich gewonnen werden. Wir versuchen das zu ändern. Auch die Datennutzer müssen von vorneherein ihre Interessen bei der Gestaltung der Fachverfahren einbringen.

Das Data Warehouse-System soll in der Endstufe das bisher eingesetzte großrechnergestützte Altsystem aller Statistikverfahren der BA ersetzen und dabei die Funktionalität und die Leistungsfähigkeit der bisherigen Plattform wesentlich übertreffen. Ein entscheidender Vorteil, den die Data Warehouse-Konzeption auch für externe Nutzer mit sich bringt, liegt in der flexiblen Datenhaltung, Datenorganisation und Datenauswertung. Beispielsweise sind damit neben den Programmen für die regelmäßige statistische Berichterstattung weitere Arbeitsmittel vorhanden, um sowohl ad-hoc-Auswertungen als auch umfangreiche systematische Daten- und Plausibilitätsprüfungen vornehmen zu können.

Als Pilotanwendung für das Data Warehouse der BA ist die Beschäftigtenstatistik ausgewählt worden. Hier wird derzeit im Probetrieb das statistische Berichtsverfahren vollständig auf Basis des Data Warehouse-Systems erstellt. Die letzten Installationen und Tests werden im nächsten Jahr abgeschlossen sein, so dass dann der Produktivbetrieb aufgenommen werden kann. In der Folge werden die übrigen statistischen Fachverfahren schrittweise auf das Data Warehouse-System umgestellt.

Die Vorarbeiten der Statistik werden dem angedachten Forschungsdatenzentrum der Bundesanstalt für Arbeit zugute kommen; sie sind sogar eine wesentliche Voraussetzung für plausible und valide Daten und deren sinnvolle Verknüpfung. Wenn ein derartiges Forschungsdatenzentrum der Bundesanstalt für Arbeit, das schrittweise zu entwickeln ist, einmal arbeitet, wird die forschungsbezogene Datenbereitstellung auf eine umfassendere und systematischere Grundlage als bisher gestellt. Die Bundesanstalt für Arbeit ist daran sehr interessiert, weil sie sich davon Hinweise für eine weitere Optimierung ihrer Fördermaßnahmen erhofft. Allerdings wäre dafür eine umfassendere Betrachtungsweise und Definition von Erfolg und Misserfolg notwendig als sie bisher in wissenschaftlichen Veröffentlichungen vorherrscht.

Alle Wünsche wird aber auch ein Forschungsdatenzentrum nicht erfüllen können. Auch in Zukunft wird die Wissenschaft eigene Datenquellen erschließen müssen, um alle denkbaren Fragen beantworten zu können. Wenn sich diese Erkenntnis durchgesetzt hat, wird möglicherweise eine neue Bescheidenheit Platz greifen.