

## **BA-Geschäftsstatistik und Forschung**

*Werner Karr, Eberhard Wiedemann*

---

### **1 Zur Historie der wissenschaftlichen Nutzung von BA-Daten für die Forschung**

Noch bis weit in die siebziger und z. T. auch in die achtziger Jahre hinein entstanden die Statistiken der Bundesanstalt für Arbeit aus Strichellisten oder im Zählblattverfahren. Diese wurden in den Arbeitsämtern erstellt und ausgewertet, in den Landesarbeitsämtern z. T. verdichtet und dann an die Hauptstelle der BA nach Nürnberg geleitet, wo daraus bundesweite Tabellen zusammengestellt wurden. Das bedeutete, dass alle Statistiken, in welcher Gliederungstiefe auch immer, aggregiert waren. Sie wurden auch fast vollständig veröffentlicht. Hausintern gab es ein Verzeichnis nicht veröffentlichter Statistiken. Hierbei handelte es sich aber nicht etwa um geheim zu haltende Zahlen, sondern um solche von eher marginaler Bedeutung, sodass von einer Veröffentlichung abgesehen wurde.

Eine wesentliche und richtungsweisende Veränderung ergab sich im Gefolge der schnell ansteigenden Arbeitslosigkeit etwa Mitte der siebziger Jahre. Seit dieser Zeit (zum ersten Mal 1973) wurden einmal jährlich so genannte Strukturhebungen der Bestände bei Arbeitslosen und offenen Stellen mit einer Fülle von Merkmalen durchgeführt. Auch dies geschah mit Hilfe von Zählblättern (pro Person oder pro Stelle), die zentral in Nürnberg in die Datenverarbeitung eingelesen wurden und zu Dateien mit Individualdatensätzen führten. Obwohl die Ergebnisse in standardisierten Tabellen präsentiert wurden, boten die Daten den Vorteil, Auswertungen unabhängig von gängigen Gliederungen oder auch in Kombination mehrerer Merkmale vornehmen zu können. Die dabei entstandenen Dateien wurden vom IAB, aber auch von anderen Instituten, zur differenzierten Auswertung genutzt. Probleme des Datenschutzes bestanden dabei nicht. Zum einen war Datenschutz noch kein Thema, zum anderen waren die Zahl und Differenzierung der Merkmale noch zu gering, um Einzelfälle identifizieren zu können.

Im Zuge der weiteren Entwicklung wurden solche Erhebungen allmählich überflüssig, da die benötigten Daten in den so genannten Fachverfahren DV-mäßig gespeichert wurden. Die Nutzung dieser so gespeicherten Daten für Forschungszwecke wurde dadurch komplizierter. Denn die dort gespeicherten Daten hatten überwiegend fachliche Bezüge, z. B. leistungsrechtlicher Art bzw. waren sie am Aufgabenvollzug der Vermittlung oder Beratung orientiert. Für statistische Zwecke mussten sie in der Regel bearbeitet und statistischen Definitionen angepasst werden.

Hinzu kam, dass die so genannten Fachverfahren meist als zwischen den Fachbereichen unverbundene Insellösungen entstanden. Dies ist für das hier behandelte Thema deswegen relevant, weil hierdurch Weichenstellungen erfolgt sind, die zu Folgeproblemen führten. Im Gegensatz beispielsweise zu den Rentenversicherern, bei denen eine Person langfristig Objekt der Datenhaltung ist (ein 20-jähriger Azubi muss mindestens 45 Jahre im Computer gespeichert bleiben, um mit 65 Jahren seine Rente zu bekommen), waren die entsprechenden Konzepte in der BA eher auf kurzfristige „Fälle“ ausgelegt. Jemand wurde arbeitslos, hoffentlich nur für kurze Zeit, in dieser Zeit wurden ihm Stellenangebote gemacht; die Zeit wurde überwiegend mit Lohnersatzleistungen überbrückt. Nach Beendigung der Arbeitslosigkeit war „der Fall“ abgeschlossen und wurde (nach kurzen Zeiten) gelöscht.

Ein entsprechendes Konzept lag auch anderen Verfahren, z. B. der Förderung der beruflichen Weiterbildung o. Ä. zugrunde. Im Datensatz des Arbeitslosen oder Bewerbers wurde bei einem solchen Übergang als Beendigung der Arbeitslosigkeit lediglich festgehalten, dass er in eine Weiterbildungsmaßnahme einmündet, in der Datei der Weiterbildungsteilnehmer war er aber mangels eines einheitlichen Identifizierers nicht ohne weiteres zu finden. Noch komplizierter war es, wenn ein Arbeitsloser in eine Arbeitsbeschaffungsmaßnahme zugewiesen wurde. Die ABM-Datei war nämlich nicht personen- oder fallbezogen, sondern maßnahmebezogen. Von einem Arbeitslosen, der eine ABM-Beschäftigung aufnahm, wusste man nur dies als Abgangsgrund. Man konnte ihn deswegen ebenfalls nicht in der Beschäftigtenstatistik finden, in die er inzwischen eingemündet ist.

Dies alles hat die effiziente Durchführung der den Fachabteilungen übertragenen Arbeiten nicht behindert. Es war in den siebziger Jahren auch nicht zu antizipieren, dass die Massenarbeitslosigkeit ein dauerhaftes Phänomen werden würde und dass beispielsweise für die erwerbsbiographische Forschung oder die Evaluation arbeitsmarktpolitischer Maßnahmen komplexe verknüpfte Individualdatensätze benötigt würden. Diese datentechnischen Probleme wurden im konzeptionellen Vorfeld der Einführung der Eingliederungsbilanzen 1999 besonders deutlich.

Mit der Entstehung eines „Data-Warehouse“ in der BA, an dessen Entwicklung das IAB seit mehreren Jahren aktiv mitwirkt, wird sich die Nutzungsmöglichkeit von Prozessdaten der BA in Zukunft deutlich verbessern - aber nur schrittweise lassen sich die Restriktionen beheben, die in der Vergangenheit sowohl konzeptionell als auch hinsichtlich der Leistungsfähigkeit von Hard- und Software der DV-Ausstattung bestanden.

## **2 Zur Entwicklung der Rechtslage**

### ***Arbeitsförderungsgesetz***

Im Arbeitsförderungsgesetz (AFG) vom 25.06.1969 wurde die schon zwei Jahre früher (1967) begonnene Arbeitsmarkt- und Berufsforschung in der Bundesanstalt für Arbeit (BA) gesetzlich verankert. Dort wurde auch festgelegt, dass die BA die für diese Forschung „erforderlichen Unterlagen zu erstellen, zu führen und auszuwerten“ hat (§ 6 Abs. 2 AFG). Damit war für das IAB eine gute Rechtsgrundlage für empirische Forschung gegeben: Das IAB konnte nicht nur alle Statistiken und sonstige Geschäftsdaten des Hauses BA nutzen, es konnte auch darüber hinaus auf Veränderungen dieser Statistiken einwirken, es konnte zusätzliche Statistiken anregen, es konnte über die Arbeitsämter flächendeckend oder stichprobenweise Sondererhebungen durchführen und hatte zudem Mittel im Forschungstitel des Haushaltes für externe Erhebungen. Alle diese Möglichkeiten wurden auch genutzt.

### ***Sozialgesetzbuch III***

Das AFG wurde am 01.01.1998 durch das SGB III abgelöst. Die zuvor geschilderten Sachverhalte wurden zum Teil anders formuliert, z. T. auch in den Zuständigkeiten anders geregelt. Die Arbeitsmarkt- und Berufsforschung der BA ist im § 282 Abs. 1 SGB III verankert. Die Datennutzung ist für IAB und Statistik getrennt geregelt: nach § 281 SGB III hat die BA „aus den in ihrem Geschäftsbereich anfallenden Daten Statistiken ...“ zu erstellen. Der Zugang des IAB zu den Daten des Hauses ist in § 282 Abs. 2 (ab 01.01.2002 Abs. 5) SGB III geregelt, dort heißt es u. a.: Innerhalb der BA dürfen die Daten aus ihrem Geschäftsbereich dem IAB zur Verfügung gestellt und dort für dessen Zwecke genutzt und verarbeitet werden. Das IAB darf ergänzende Erhebungen ohne Auskunftspflicht der zu Befragenden durchführen, wenn sich die Informationen nicht bereits aus den im Geschäftsbereich der BA vorhandenen Daten oder aus anderen statistischen Quellen gewinnen lassen.

In § 282 Abs. 3 (ab 01.01.2002 Abs. 6) wird dem IAB schließlich die Führung der so genannten Historik-Datei, also die Aufbewahrung aller beschäftigungsrelevanten Meldungen über sozialversicherungspflichtig Beschäftigte seit Beginn des Meldeverfahrens 1975, neuerdings auch der geringfügig Beschäftigten, zugewiesen. Dies erfolgte deshalb, weil für Forschungszwecke Daten in so langen Zeiträumen benötigt werden, wogegen sie für Verwaltungszwecke nicht so lange aufbewahrt werden dürfen.

Das IAB hat damit nach wie vor Zugang zu allen Daten der BA; diese können entsprechend dem Bedarf für Forschungszwecke beim IAB zusätzlich länger gespeichert werden als im Rahmen der normalen Aufgabenabwicklung der Bundesanstalt für Arbeit zulässig.

Rechtlich sind die Änderungen bezüglich des Datenzugangs relativ gering; sie waren eher präzisierend.

### ***Job-AQTIV-Gesetz***

Durch das Job-AQTIV-Gesetz wurde ab 01.01.2002 das SGB III novelliert. Im bestehenden § 282 SGB III wurden, aufsetzend auf den bisherigen Regelungen, die Anforderungen an die Arbeitsmarkt- und Berufsforschung erweitert. Nach dem neuen Absatz 2 ist die zeitnahe Untersuchung von Wirkungen der Arbeitsförderung (Wirkungsforschung) nunmehr ein Schwerpunkt der Arbeitsmarktforschung und richtet sich insoweit an jede Forschung zum Arbeitsmarkt. Für das IAB wurde Wirkungsforschung als eine ständige Aufgabe festgeschrieben, ein Schwerpunkt neben der Grundlagenforschung und dem breiten Spektrum analytischer Arbeiten.

Ausführliche Anforderungen an die Wirkungsforschung werden im Absatz 3 vorgegeben. Demnach hat Wirkungsforschung von den - durchaus komplexen - Zielen der Arbeitsförderung auszugehen, die mit dem Job-AQTIV-Gesetz im § 1 SGB III neu definiert wurden. Gefordert wird, dass neben der primären Zielsetzung der Integration in reguläre Beschäftigung, auch Auswirkungen auf die individuelle Beschäftigungsfähigkeit untersucht werden sowie Aussagen über Kosten-Nutzen-Relationen bis hin zu gesamtwirtschaftlichen Nettoeffekten getroffen werden. Durch Analysen von Erwerbsverläufen sollen langfristige Wirkungen festgestellt werden, auch im Hinblick auf die Chancengleichheit der Geschlechter. Von der Forschung werden belastbare Ergebnisse für die regionale Ebene, also auch für Arbeitsämter oder andere Regionszuschnitte, erwartet.

Der neue Absatz 7 des § 282 SGB III soll den Zugang zu BA-Daten für externe Forscher erleichtern. Erstmals wird es für die BA zur gesetzlichen Pflicht, wissenschaftlichen Einrichtungen auf Ersuchen anonymisierte Daten für Zwecke der Arbeitsmarkt- und Berufsforschung zur Verfügung zu stellen. Letztendlich sollen die Festlegungen in den Absätzen 2 und 7 den Wettbewerb in der wissenschaftlichen Arbeitsmarktforschung grundsätzlich ermöglichen und befördern. Aus dem Job-AQTIV-Gesetz erwachsen an die BA zusätzliche Anforderungen für Datenbereitstellung und Datenhaltung für Forschungszwecke.

### ***Aspekte des Datenschutzes***

Seit dem „Volkszählungsurteil“ vom 15.12.1983, dem Bundesdatenschutzgesetz (BDSG) vom 27.01.1977 und dem für Daten der BA maßgeblichen SGB X vom 01.01.1981 sind personenbezogene Daten geschützt. Dies trifft zusammen mit einer Entwicklung, bei welcher pro Person immer mehr und für eine Reidentifikation immer wichtigere Daten (z. B. datierte Zeiträume) gespeichert werden, sodass im Allgemeinen davon auszugehen

ist, dass inzwischen vorliegende Individualdatensätze über Arbeitslose, Beschäftigte, Maßnahmeteilnehmer o. Ä. nicht anonym sind. Seit zu prüfen ist, ob Daten als faktisch anonym gelten können oder ob nicht anonymisierte Daten nach vorheriger Genehmigung durch die oberste Bundesbehörde übermittelt werden dürfen stellt sich auch die Frage, wer in der BA darüber zu entscheiden hat. Im Zusammenhang damit hat sich die Redewendung vom „Herr der Daten“ herausgebildet: Vielfach ist dies die Abteilung Statistik; in einigen Fällen (z. B. Beschäftigtenhistorikdatei) gibt es eine eigene Zuständigkeit des IAB, dieser Anteil wird sich künftig durch die Generierung spezieller Datensätze erhöhen.

### **3 Strategien für Verbesserung der Datenlage für Forschungszwecke**

#### ***Grundsätzliches***

Arbeitsmarkt- und Berufsforschung hat viele Interessen zu bedienen: vielfältige Informationsbedürfnisse der Politik, der Arbeitsverwaltung, der Sozialpartner und der Öffentlichkeit. Gleichzeitig erstrebt sie eigenen Erkenntnisgewinn. Arbeitsmarktforschung, auch im IAB, war und ist geprägt durch ein breites Spektrum analytischer Arbeiten zur Arbeitsmarktentwicklung, zur Nachfrage der Betriebe nach Arbeitskräften, zur Entwicklung von Berufen und Qualifikationen und vieles mehr. Die eingangs beschriebene Datenbasis der BA, die Geschäftsstatistik, mit aggregierten oder teilaggregierten Daten in z. T. tiefer Gliederung und in Verbindung mit anderen Statistiken z. B. des Statistischen Bundesamtes oder der Sozialversicherungsträger genügten im Wesentlichen den Forschungsbedürfnissen. Als überwiegend unzureichend erwiesen sich die verfügbaren Daten jedoch für anspruchsvolle Evaluationsfragestellungen. Ganz neue Anforderungen an Art und Umfang der Daten entstehen allein schon daraus, wenn man z. B. mit methodisch weit entwickelten (mikro)ökonomischen Modellen den Erfolg eines bestimmten arbeitsmarktpolitischen Instruments oder eines Maßnahmebündels nachweisen will. Die Heterogenität der Teilnehmer und Maßnahmen ist möglichst genau abzubilden, die Beobachtungsdauer bis weit nach Maßnahmeende auszudehnen. Wirkungsforschung müsste mit ökonomischen Experimenten einhergehen, was im Ausland bereits seit längerem geschieht, in Deutschland aber eher als nicht machbar erscheint. So greift man zum Quasi-Experiment: statistische Zwillinge für Nichtteilnehmer an einer Maßnahme finden und als Vergleichsgruppe (Kontrollgruppe) nutzen. Gleiches gilt aber auch für den regionalen Kontext, um den Einfluss der Verfassung regionaler Arbeitsmärkte als Erfolgs- oder Misserfolgsindikator für arbeitsmarktpolitische Maßnahmen zu erfassen. Leistbar wird dies alles nur, wenn Individualdatensätze mit einem möglichst breiten Merkmalspektrum vorhanden sind. In den Fachverfahren der BA entstehen sie, sie sind nur bisher nicht, bzw. nur in Ausnahmefällen, auswertungsreif (zentral) verfügbar und untereinander, z. B. Personendaten und Maßnahmedaten, nicht einfach kombinierbar.

Weiteres ist zu bedenken: über die im Gesetz festgeschriebenen Anforderungen hinaus erwartet die Politik (der Gesetzgeber) auch die Erfolgskontrolle von Maßnahmen, die von Dritten im Auftrag der BA durchgeführt werden. Gleiches gilt für Modellprojekte der Arbeitsmarktpolitik, wo insbesondere eine zeitnahe und kontinuierliche Begleitforschung erfolgen soll. Wirkungsforschung zum arbeitsmarktpolitischen Instrumentarium soll auch belastbare, empirisch abgesicherte Ergebnisse für eine bessere Steuerung der Arbeitsmarktpolitik liefern. Dies gilt nicht nur für die Bundesebene, sondern gleichermaßen auch für den Zielsteuerungsprozess in der BA selbst. Es besteht also ein erhebliches Eigeninteresse an entsprechend strukturierten Daten, die valide Evaluationsforschung gewährleisten. Darüber hinaus sind Querverbindungen zu sehen und datenseitig abzusichern: zum BA-internen Controlling, zur (noch aufzubauenden) Kostenrechnung und zu Effizienzbetrachtungen.

Nicht zuletzt beginnt Datenbereitstellung mit der Datenerfassung vor Ort, in den Arbeitsämtern, bei der einzelnen Fachkraft. Deren Interesse ist, für den „Kunden“ mit seinem Anliegen (z. B. Vermittlung in Arbeit, Leistungsgewährung) möglichst viel Zeit zu haben, aber wenig Zeit aufwenden zu müssen für Datenerfassungen, deren Nützlichkeit sich ihr nicht vordergründig erschließt. Den Fachkräften vor Ort muss aber bewusst werden, dass sie letztendlich mit ihren Erfassungen eine amtliche Statistik erzeugen, sie den Grundstein für valide Forschungsergebnisse legen.

### ***Kooperationsprojekte, Maßnahme-Teilnehmer-Grunddatei***

Die Ausführungen zeigen, dass damit ein statistischer Neubeginn für diese Art von Datensätzen erforderlich wurde, der nach längerem konzeptionellen Vorlauf im Jahr 2000 eingesetzt hat und bereits vieles in Bewegung brachte bis hin zu neuen informationstechnischen Infrastrukturen in der BA und Einkauf externen Sachverständigen. Die Komplexität des Gesamtanliegens und die Kompliziertheit der einzelnen Teilschritte erlauben nur ein abgestuftes Vorgehen mit einem zeitlich in weiter Ferne liegenden Realisierungshorizont für das (vorläufige) Endziel Data-Warehouse. Soviel Zeit bleibt nicht, die Wissenschaft fordert vehement evaluationstaugliche Datensätze, die Politik erwartet bald Antworten auf ihre drängenden Fragen. Das IAB nimmt sich dem im Besonderen an, schafft Vorlauf und Lösungen für das in der BA umzusetzende Datenkonzept. Es baut dabei u. a. auf Kooperationsprojekten mit der externen Wissenschaft, um Wissenstransfer zu nutzen und zusätzliche Kapazitäten für Entwicklungsarbeiten zu erschließen.

Im Herbst 2000 hat das IAB mit einem neuen Projekt die Evaluation der Beschäftigungswirkungen von Fortbildungs- und Umschulungsmaßnahmen (FuU) mit prozessproduzierten Daten begonnen in Kooperation mit Lehrstühlen der Universitäten Mannheim und St. Gallen. Das Forschungsdesign erfordert Daten, die der Heterogenität von Maß-

nahmen, Teilnehmern und Regionen angemessen Rechnung tragen, Maßnahmekarrieren erkennen lassen und die zeitliche Verortung der Maßnahme in der jeweiligen Erwerbsbiographie ermöglichen. Dies ist nur mit der Generierung eines neuartigen Datensatzes realisierbar. Hierzu wird die IAB-Beschäftigtenstichprobe von 1980-1997 verbunden mit der BA-Statistik über die Teilnehmer an FuU-Maßnahmen (sog. St 35, erst ab 1980 verfügbar) sowie abgeglichen und ergänzt mit Daten der Leistungsempfängerdatei (Nachweis arbeitslos) und Meldungen der Beschäftigtenstatistik (Nachweis sozialversicherungspflichtig beschäftigt). Die Datenfusion war wesentlich schwieriger und zeitaufwändiger als erwartbar (z. B. Informations- und Dokumentationslücken, Inkonsistenzen in der zeitlichen Struktur), viele Probleme konnten nur durch intensive Einzelfallbetrachtung „per Hand“ gelöst werden. Im Jahr 2002 wird der konsolidierte Datensatz nutzbar sein, die Evaluation beginnen können, die 2005 abgeschlossen werden soll.

Ein weiteres Projekt zur Evaluation von Arbeitsbeschaffungs- und Strukturanpassungsmaßnahmen (ABM und SAM) hat das IAB im Frühjahr 2001 in Kooperation mit dem Lehrstuhl für Statistik und Ökonometrie der Universität Frankfurt/M. gestartet, das in einer ersten Etappe bis Ende 2004 reicht und Anfang 2006 abgeschlossen sein wird. Unter Anwendung ökonometrischer Verfahren sollen die durchschnittlichen Nettoeffekte für Teilnehmer an ABM/SAM-Maßnahmen festgestellt werden. Für die Datengewinnung wird ein neuer Ansatz verfolgt: Grundlage sollen prozessgenerierte Datensätze der BA sein, es werden also die Daten aus den Fachverfahren unmittelbar genutzt. Grundidee ist, am Beispiel dieses Forschungsvorhabens eine „Maßnahme-Teilnehmer-Grunddatei“ systematisch aufzubauen. Dabei wird ein Datenmodell entwickelt, das nicht nur die Teilnahme an ABM- und SAM-Maßnahmen abbilden kann, sondern auch die anderen Instrumente der aktiven Arbeitsförderung mit einschließt: Förderung der beruflichen Weiterbildung (FbW), Trainingsmaßnahmen, Lohnkostenzuschüsse, Überbrückungsgeld, Freie Förderung (FF) u. a. In der aufzubauenden Datenbank werden einerseits maßnahmebezogene, andererseits aber auch erstmals personenbezogene Recherchen möglich sein. Auf diese Weise entsteht eine Datenbank, die es erlaubt, „Maßnahme-Karrieren“ zu verfolgen. Zudem gestattet der in dieser Datenbank integrierte Personenstamm die Verknüpfung mit allen wesentlichen Datenquellen der BA: mit der Bewerberangebotsdatei der Arbeitslosen und der nicht arbeitslos Arbeitssuchenden (BewA), mit den Daten der Beschäftigtenstatistik (BST) und mit der Leistungsempfängerdatei (LED). Somit wird in Zukunft eine übergreifende „Karriere“-Verfolgung möglich (Beschäftigung, Arbeitslosigkeit, Weiterbildung und andere Instrumente der Arbeitsförderung). Zudem wird versucht, unter den gegebenen Bedingungen eine optimale Datengrundlage für die methodisch aufwändige Vergleichsgruppenbildung und für eine Verbleibsanalyse bis zum aktuellen Rand aufzubauen. Dieses anzugehen war erst möglich, nachdem in 2000 flächendeckend, aber noch nicht für alle Verfahren und Statistiken, als Identifikator die Kunden-

nummer eingeführt wurde. Über „Brückenkonstruktionen“ lassen sich jedoch noch fehlende direkte Verknüpfungsmöglichkeiten überwinden. Das IAB nutzt die neue Datenbasis in einer Reihe weiterer Evaluationsprojekte, z. B. für die Begleitforschung zum ESF-BA-Programm 2000-2006. Unter Beachtung des Datenschutzes stehen die extrahierten Grunddaten selbstverständlich auch externer Forschung zur Verfügung.

Die Datenbereitstellung im Rahmen des ABA/SAM-Kooperationsprojektes ist nur der erste, aber entscheidende Schritt zum Aufbau einer allgemeinen und dauerhaft verfügbaren auswertungsbereiten Datenbasis für Evaluationszwecke. Aufbauend auf diesen Vorarbeiten des IAB sollen die Ausgangsdaten für eine derartige Datenbasis künftig über das Data-Warehouse der BA bereitgestellt werden. Spezielle für Forschungszwecke aufzubereitende Daten, z. B. Erwerbskarrieren, sollen im plattformübergreifenden allgemeinen Analyse- und Informationssystem „pallas“ des IAB aktualisiert und gepflegt werden.

### ***Pallas und Data-Warehouse***

Das **plattformübergreifende allgemeine Analyse- und Informationssystem - pallas** - des IAB befindet sich im Aufbau. Es wird ausgehend von der BA-Datenbasis spezifische Forschungsdaten für generelle Fragestellungen der Arbeitsmarkt- und Berufsforschung oder auch für konkrete Projekte aufbereiten und bereitstellen. Neben klassischer Software der statistischen Analyse (z. B. SAS, SPSS) werden auch Data-Warehouse-Methoden und Technologien eingesetzt. Sie gestatten u. a. eine Verknüpfbarkeit von Daten aus unterschiedlichsten Quellen (konsolidierte Daten aus dem künftigen BA-DWH bzw. Basisdaten aus den operativen Altverfahren, Fremddaten, IAB-eigene Daten), automatisierte Datenaktualisierungen und komfortable multidimensionale Online-Analysen.

Die für die aktuellen Bedarfe der Wirkungsforschung erforderlichen Daten können derzeit nur aus den bestehenden Altverfahren gewonnen werden. Sie werden mit Daten aus anderen Quellen zu einer konsolidierten verknüpfbaren Datenschicht zusammengeführt. Diese speziell für Forschungszwecke konsolidierte Datenbank bildet das Kernstück des IAB-Analyse- und Informationssystems. In Form multidimensionaler Würfel werden die Daten für unterschiedlichste Verwertungskontexte online analysierbar sein. Die mit pallas entwickelten Modellierungs-, Prüf- und Bereinigungsregeln werden später als Anforderung der Forschung an das BA-Data-Warehouse übergeben und umgesetzt, soweit es sich um allgemein relevante Aufbereitungsschritte handelt. Die nur für die Forschung spezifischen Aufbereitungen werden auch in Zukunft in pallas erfolgen.

Die BA hat sich zum Ziel gesetzt, mit dem Aufbau eines Data-Warehouse (DWH) die Qualität und die Auswertungsmöglichkeiten der geschäftspolitisch relevanten (dispositiven) Daten nachhaltig zu verbessern und einem breiten Nutzerkreis zugänglich zu ma-



chen. Auch beim DWH der BA handelt es sich um eine zentrale Datenbasis, in die Daten aus unterschiedlichen operativen internen IT-Verfahren (z. B. Arbeitsvermittlung, Ausbildungsstellenvermittlung, Leistungsgewährung, Orts- und Betriebsdatenverwaltung), verschiedenen Geschäftsfeldern (z. B. Finanzen, Controlling) aber auch von externen Stellen (z. B. Sozialversicherungsträgern) eingestellt, geprüft, aufeinander abgestimmt und zu einer einheitlichen Datenbasis zusammengefasst werden. Im DWH werden auch zur Erfüllung der fachlichen Aufgaben und für Forschungszwecke notwendigen Individualdaten in datenschutzgerechter Form gespeichert. Es werden DWH-technisch Vorkehrungen getroffen, dass die variablen Auswertungserfordernisse der unterschiedlichsten Nutzer (Verwaltung, Statistik, Forschung) mit hoher Performance realisiert werden. Die Auswertungswerkzeuge sollen komfortabel sein und Online-Zugänge gewährt werden. Die Konfigurierung des DWH ist nach einem längeren Findungsprozess festgelegt, Hardware bereits installiert, die erforderliche Software für Betrieb, Datenbanksystem, multidimensionale Datenhaltung und Auswertungswerkzeuge ausgewählt und beschafft. Im Jahr 2002 wird ein DWH-Pilotsystem in Betrieb sein und mit Echtdateien arbeiten. Als erste Datenbasis wird die Beschäftigtenstatistik, beginnend mit dem Jahr 2000, mit allen zugehörigen Dimensionsdaten (Gebietsstrukturdaten, Betriebsdaten, Wirtschaftsklassifikationen u. a.) auswertungsreif verfügbar sein. Die DWH-Datenbasis wird ab 2002 schrittweise erweitert, zuerst mit den statistischen Bewerber- und Stellenangebotsdaten und danach mit den Maßnahmedaten. Parallel zum Aufbau der zentralen Datenbasis werden die Fachverfahren den neuen Anforderungen angepasst bzw. neu aufgebaut. Das Projekt Data-Warehouse der BA ist auf den Weg gebracht, mit guten Perspektiven auch für potenzielle Nutzer aus dem Wissenschaftsbereich, noch offen in der zeitlichen Dimension, bis wann welche Daten der Öffentlichkeit tatsächlich der Auswertung zugänglich sind.

### ***Forschungsdatenzentrum***

Die Datenquellen der BA zählen zu den wichtigen nationalen Datenressourcen, sie ist wie geschildert damit ein wichtiger Datenproduzent für Zwecke der wissenschaftlichen Forschung. Nach den Vorstellungen der vom Forschungsministerium einberufenen Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik (KVI), des Forschungsministeriums (BMBF) selbst und des im gleichen Zusammenhang vom BMBF gebildeten Gründungsausschusses des Rates für Sozial- und Wirtschaftsdaten sollen Forschungsdatenzentren (FDZ) an mehreren Standorten entstehen, namentlich und vorrangig beim Statistischen Bundesamt und bei der Bundesanstalt für Arbeit. Entsprechend der von der KVI formulierten Vorstellungen bestehen die grundlegenden Ziele des Forschungsdatenzentrums der BA in der Bereitstellung und projektspezifischen Aufbereitung von nicht anonymisierten bzw. nichtanonymisierbaren

Einzeldaten für die Forschung. Aus diesem Grunde ist das FDZ gegenüber der Exekutive und anderen Unbefugten besonders abzuschotten.

Vom Statistikbereich der BA und vom IAB sind mit den geschilderten Aktivitäten Grundlagen geschaffen worden, die für das FDZ nutzbringend verwertet werden können. In der konkreten Ausgestaltung des FDZ sollen - für das gesamte Angebot an relevanten Einzeldaten - die in der Praxis bereits bewährten Funktionen der IAB-Schalterstelle zum IAB-Betriebspanel übernommen und erweitert bzw. verbessert werden. Das heißt primär, dass externe Wissenschaftler die im FDZ verfügbaren Daten in schwach anonymisierter Form (z. B. ohne Versicherungsnummer, Namen, Adresse) oder, sofern die Fragestellung dies erforderlich macht, in nicht anonymisierter Form unter Berücksichtigung der geltenden Datenschutzbestimmungen persönlich innerhalb des FDZ auswerten können. Das FDZ kann außerdem als Schalterstelle ablauffähige Programme der externen Wissenschaftler ausführen und die Auswertungsergebnisse nach Prüfung der datenschutzrechtlichen Zulässigkeit zurücksenden (kontrollierte Datenfernverarbeitung). Denkbar sind auch (kostenpflichtige) Sonderauswertungen, die Erstellung von Scientific Use Files (wie vom IAB bereits geleistet) oder für bestimmte Auswertungsbedarfe ein Web-gestütztes Online Analytical Processing, das einen Zugriff über geschützte Leitungen auf einen stark disaggregierten Datenbestand zulässt, ohne direkt mit den schützenswerten Einzeldaten in Berührung zu kommen. Ein dementsprechender Pilotversuch soll in 2002 vom IAB gestartet werden.

Um eine effektive und qualifizierte Verwendung der Datensätze durch die externe Forschung zu gewährleisten, müssen die Aufgaben des FDZ auch Dokumentationsarbeiten sowie Benutzerberatungen zu Daten, Methoden und Einsatz von Auswertungswerkzeugen beinhalten. Darüber hinaus sind eigene methodische Forschungen zur Sicherung der Qualität und Verwendbarkeit der Daten zu leisten, soweit dies für die auftragsspezifische Datenaufbereitung und die Durchführung von Anonymisierungsmaßnahmen erforderlich ist.

Auf verschiedenen Ebenen und in unterschiedlichen Zusammenhängen wird die Datenbasis für Forschungszwecke forciert entwickelt und verbessert. Dies nährt die Hoffnung, in absehbarer Zeit die aus dem Ausland wohlbekannten Untersuchungen zur Wirkung arbeitsmarktpolitischer Hilfen mit eigener, deutlich verbesserter Datenbasis durchführen zu können. Das IAB und die BA werden sich deshalb auch weiterhin dafür einsetzen, dass externe Forscher mit den neu entstehenden Datensätzen uneingeschränkt arbeiten können.