

**Pilotprojekt  
„Beschleunigte Datenbereitstellung  
für externe Institute zum Zwecke der Evaluation  
arbeitsmarktpolitischer Instrumente“**

Christian Brinkmann (IAB)



## Pilotprojekt „Beschleunigte Datenbereitstellung für externe Institute zum Zwecke der Evaluation arbeitsmarktpolitischer Instrumente“<sup>1</sup>

### 1. Hintergrund

Evaluationsforschung hat in allen Politikbereichen und in der Wissenschaft ein zunehmendes Gewicht erhalten.<sup>2</sup> Angesichts der langanhaltenden Massenarbeitslosigkeit gilt dies namentlich für die Arbeitsmarktpolitik mit ihrem immensen Mittelaufwand.<sup>3</sup> Mit ihr sollen vor allem Arbeitsmarktchancen von Benachteiligten verbessert werden, es geht um den Erhalt von Beschäftigungsfähigkeit und die Vermeidung sozialer Exklusion – so auch entsprechende Zielsetzungen auf EU-Ebene. Der Strukturwandel führt in den Industriestaaten z.B. dazu, dass immer weniger Arbeitsplätze für Geringqualifizierte angeboten werden und Ältere langzeitarbeitslos werden bzw. vorzeitig aus dem Erwerbsleben ausscheiden, weil es nicht gelingt, ihre Kenntnisse und Fertigkeiten im Sinne eines lebenslangen Lernens zu erhalten bzw. neuen Anforderungen anzupassen.<sup>4</sup> Mit den Instrumenten der Arbeitsmarktpolitik können in diesem Zusammenhang unverzichtbare Beiträge zum Erhalt und zur Weiterentwicklung des Humankapitals geleistet werden.

Ob und inwieweit diese Ziele für welche Personengruppen in welcher Konstellation aber auch tatsächlich erreicht und die Beschäftigungschancen tatsächlich verbessert werden, war bislang wissenschaftlichen Untersuchungen (zumindest auf der repräsentativen Ebene) kaum zugänglich. Ganz abgesehen von dem sich jüngst erst entwickelnden Evaluationsbewusstsein: Es fehlte an den für Evaluationszwecke unverzichtbaren *Individualdaten* mit hinreichend differenzierten Informationen über die (in sich sehr heterogenen) Maßnahmen, über die Struktur der Teilnehmer/innen selbst und über die jeweiligen (vor allem regionalen, arbeitsmarktbezogenen) Kontexte des Maßnahmeeinsatzes.

Hier ist nun grundsätzlich ein entscheidender Wandel realisierbar. Hard- und Software-Entwicklungen im DV-Bereich machen es nunmehr möglich, dass die entsprechenden Prozessdaten der BA der Wissenschaft zugänglich gemacht werden. Diese Entwicklung soll mit dem vorliegenden Projekt ganz im Sinne des KVI-Gutachtens an entscheidender Stelle unterstützt werden.

---

<sup>1</sup> Bezug genommen wird auf das Gutachten der Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik (KVI) mit der Empfehlung zur Einrichtung von Forschungsdatenzentren und Empfehlungen des Gründungsausschusses des Rates für Sozial- und Wirtschaftsdaten, im Vorfeld hiervon Pilotprojekte, die vorrangig der Erschließung von Datensätzen dienen, zu vergeben.

<sup>2</sup> Vgl. Reinhard Stockmann (Hrsg. 2000): Evaluationsforschung. Grundlagen und ausgewählte Forschungsfelder, Leske und Budrich, Opladen

Reinhard Hujer, Lutz Bellmann, Christian Brinkmann (Hrsg. 2000): Erfolgskontrolle der Arbeitsmarktpolitik, Schwerpunkt-Heft der Mitteilungen aus der Arbeitsmarkt- und Berufsforschung, Heft 3

<sup>3</sup> Allein über den Eingliederungstitel der BA werden in diesem Jahr 27,8 Mrd. DM aufgewendet. Hinzu kommen weitere Mittel von Bund, Länder, Kommunen und EU.

<sup>4</sup> Vgl. Christian Brinkmann, Karen Schober (Hrsg. 1992): Erwerbsarbeit und Arbeitslosigkeit im Zeichen des Strukturwandels, BeitrAB 163, Nürnberg

## 2. Ziele des Pilotprojekts

Im Vorfeld geht es um eine *Beschleunigung der Aufbereitung und Dokumentation* von Ausgangsdatensätzen über Maßnahmeteilnehmer, die beginnend mit dem 1.1.2000 in einer Vorstufe des entstehenden „Data Warehouse“ der BA flächendeckend und kumulativ so gehalten werden, dass über Identifizierer (Kundennummern) grundsätzliche Verknüpfbarkeit mit weiteren Prozessdaten der BA gewährleistet ist. Erschlossen werden können damit z.B. Name, Adresse und Versicherungsnummer für Abgleiche mit der Beschäftigtendatei und/oder für Befragungen. Möglich wird damit auch die Bildung von für Evaluationen unverzichtbaren Kontrollgruppen (z.B. mit ökonomischen „matching“-Verfahren auf der Basis der Arbeitslosen- oder Beschäftigtendatei mit all den dort auch für Maßnahmeteilnehmer gespeicherten Merkmalen). Die beschleunigt aufzubauende Teilnehmerdatei ist notwendige, aber keineswegs hinreichende Voraussetzung für Wirkungsforschung.

Darauf aufbauend soll konkret erprobt werden, ob und inwieweit es möglich ist, nach den gängigen Anonymisierungsregeln ein für unterschiedliche Evaluationsstrategien verwendbares Scientific Use File mit vertretbarem Aufwand zu erstellen. Im Rahmen einer solchen „Machbarkeitsstudie“ geht es um die Erstellung eines oder mehrerer „Prototypen“. Einzuschätzen wird es sein, ob und inwieweit die notwendigen Vergrößerungen (Maßnahmearten, Regionen, Maßnahmekarrieren) und das Entfernen von Identifizierern (notwendig für Befragungen und spätere Datenabgleiche) noch fundierte Evaluationen ermöglichen – mit Berücksichtigung der Heterogenität von Teilnehmern, Maßnahmen und Zielsetzungen. „Die entscheidende Grenze für die Nützlichkeit von Scientific Use Files sind Daten-files, bei denen schon die faktische Anonymisierung die Aussagekraft der Daten in hohem Maße einschränkt.“ (KVI-Gutachten S. 32).

Das Ergebnis der Machbarkeitsstudie wird nicht nur unter Evaluationsgesichtspunkten zu bewerten sein: Faktisch anonymisierte Individualdaten der Teilnehmer von Fördermaßnahmen und deren Weitergabe an externe Forscher können auch unabhängig hiervon ein wesentlicher Fortschritt sein, wenn dadurch z.B. neue Zugänge für erwerbsbiografische Fragestellungen eröffnet werden.

Es ist davon auszugehen, dass die Anonymisierung komplexer Daten einen Zeitaufwand erfordert, der inhaltliche Analysen für eine gewisse Zeit, die nicht vorab bestimmbar ist, praktisch ausschließt. Deshalb wird es notwendig sein, externen Wissenschaftlern bereits im Vorfeld eines bereinigten und voll dokumentierten Scientific Use Files Erfahrungen beim Umgang mit den benötigten Prozessdaten zu vermitteln, weil diese von der Datenhaltung her zunächst grundsätzlich nur bausteinartig (wenn auch verknüpfbar) erschlossen werden können und weiterer Bearbeitung bedürfen. Auch wenn die Anonymisierung geschützter Sozialdaten grundsätzlich Vorrang hat, wird es im Hinblick auf den jeweiligen Forschungszweck und die anzuwendenden

Methoden im Evaluationszusammenhang zunächst auch darum gehen müssen, nicht anonymisierte Datensätze projektspezifisch weiter aufzubereiten und zu verknüpfen. Dies setzt eine Genehmigung nach § 75 SGB X und eben auch besondere Kenntnisse voraus, die externen Forschern i.d.R. zunächst vermittelt werden müssen.

Gewissermaßen als Erschwernis ist bei allem zu berücksichtigen, dass die Datenaufbereitung und -haltung im Rahmen des sich entwickelnden Data Warehouse der BA für verschiedene Evaluationsstrategien (und weitere Forschungsanliegen) offen sein muß, so dass zum Beispiel auch Kontrollgruppen nicht auf Vorrat gebildet werden können.

Unterschiedliche methodische und inhaltliche Vorgehensweisen müssen im vorliegenden Zusammenhang auch deshalb möglich sein, weil sich das Erkenntnisinteresse von Wirkungsforschung nicht auf eine (für Steuerungszwecke häufig im Vordergrund stehende) kurzfristige Eingliederungszielsetzung reduzieren lässt. In dem Maße, in dem die angesprochenen Veränderungen im Erwerbssystem für viele auch mehrfache „spells“ von Arbeitslosigkeit und u.U. mehrfache Beteiligung an arbeitsmarktpolitischen Maßnahmen zur Folge haben, sind auch sich damit verbindende längerfristige Auswirkungen auf die Erwerbsbeteiligung, die berufliche Mobilität und eben „Beschäftigungsfähigkeit“ zu untersuchen. Die Zunahme von „Übergangsarbeitsmärkten“ macht es im Rahmen von berufsbiografischen Analysen geradezu unabdingbar, Wirkungen der arbeitsmarktpolitischen Interventionen in die Betrachtung einzubeziehen.

Solche Analysepotentiale gehen über die im Job-AQTIV-Gesetz geforderte Verstärkung von Wirkungsforschung noch deutlich hinaus. Wir halten die mit dem vorliegenden Pilotprojekt mögliche Beschleunigung bei der Datenerschließung einschließlich Machbarkeitsstudie hinsichtlich der Erstellung von Scientific Use Files - und soweit notwendig ergänzend Hilfestellung bei der (für viele Forscher erstmaligen) Nutzung der für Evaluationen benötigten Daten - für ausgesprochen vordringlich.

In diesem Sinne hat auch der Wissenschaftliche Beirat des IAB, dem zwölf namhafte Wissenschaftler angehören, auf seiner Sitzung am 10. Oktober 2001 die Absicht der BA, ein Forschungsdatenzentrum entsprechend dem KVI-Gutachten aufzubauen, wie auch das hier skizzierte und beim BMBF zu beantragende Pilotprojekt, ausdrücklich unterstützt.

### 3. Datenbasis

Gegenwärtig wird im IAB eine sog. Maßnahme-Teilnehmer-Grunddatei (MTG) als Teil des allgemeinen statistischen Analyse- und Informationssystem *pallas*-IAB aufgebaut. Dabei wird ein Datenmodell entwickelt, das nicht nur die Teilnahme an Arbeitsbeschaffungsmaßnahmen (ABM) und Strukturanpassungsmaßnahmen (SAM) abbildet, 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 arbeitslosen Arbeitssuchenden (BewA), mit den Daten der Beschäftigtenstatistik (BST) und mit der Leistungsempfängerdatei (LED). Somit wird in Zukunft eine übergreifende Analyse von Erwerbsbiografien ermöglicht (Beschäftigung, Arbeitslosigkeit, Weiterbildung und andere Instrumente der Arbeitsförderung).

Zur Umsetzung des oben beschriebenen Vorhabens sind komplexe fachliche und datentechnische Grundlagenarbeiten erforderlich. Wesentliche Arbeitsschritte sind dabei vom IAB bereits durchgeführt worden, weitere Arbeitsschritte stehen an und sollen mit diesem Pilotprojekt beschleunigt werden. Für die Maßnahme- und Teilnehmerdaten ABM/SAM, die eine 1. Stufe der Evaluations-Grunddatei bilden, sind dies bzw. werden dies sein:

- Konzeption und Umsetzung eines Datenmodells, das dem Anspruch genügt, alle vorhandenen Informationen zu berücksichtigen, einschließlich eines komplexen Historisierungskonzeptes;
- Konzeption und Umsetzung von Regeln und Verfahren zum Aufbau der Ladeschicht einschließlich Fehlerbehandlung: zu prüfen und zu bereinigen sind beispielsweise doppelte (bzw. mehrfache) Datensätze, ungültige Regionaleinheiten sowie ungültige Merkmalsausprägungen;
- Konzeption und Umsetzung von Regeln und Verfahren zum Aufbau der konsolidierten Datenschicht einschließlich Aufbau eines „BA-Kunden“ sowie Konsolidierung zwischen maßnahme- und teilnehmerspezifischen Daten.

Vorarbeiten in Form der Entwicklung des integrierten Datenmodells sowie in der Entwicklung und Umsetzung der Ladeschicht wurden bereits auch für weitere Datenquellen geleistet (Förderung der beruflichen Weiterbildung, Freie Förderung, Lohnkostenzuschüsse und Überbrückungsgeld). Fernziel ist die Einbeziehung der gesamten Arbeitsförderung der BA einschließlich Vermittlung und Beratung.

Beim sukzessiven Aufbau von Scientific Use Files im Rahmen dieser Machbarkeitsstudie geht es in einem ersten Schritt darum, für die bereits erschlossenen (in konsolidierter und dokumentierter Form in die Grunddatei eingestellten) Maßnahmebereiche eine erste und unverzichtbare Verknüpfung herzustellen, und zwar mit der Arbeitslosendatei (bzw. darüber hinausgehend: der Datei der Arbeitsuchenden). Ohne das damit für (fast) alle Maßnahmeteilnehmer erschließbare erweiterte Merkmalspektrum wäre das entsprechende Scientific Use File allgemein nur von geringem Nutzen und speziell für Evaluationen kaum brauchbar.<sup>5</sup> Die bisherigen Erfahrungen zeigen, dass solch ein Abgleich mit begrenztem Aufwand möglich ist, und zwar auch schon im Vorfeld der Einstellung dieser Daten in das Data Warehouse.

In einem zweiten Schritt werden aus der Datei der Arbeitsuchenden Vergleichsstichproben gezogen (potentielle Kontrollgruppen) und in gleicher Weise faktisch anonymisiert.

Hinzugefügt werden beiden Dateien (vor Entfernung der Identifizierer) „Zielvariablen“ zum Verbleib. Im Hinblick auf die gegenwärtigen Möglichkeiten heißt dies: Monatliche Angaben zum Verbleib arbeitslos oder in einer (erneuten) Maßnahme für einen noch zu bestimmenden Zeitraum (z.B. bis zu einem Jahr nach Ausscheiden aus der Maßnahme bzw. für die Vergleichsgruppe nach dem jeweiligen Referenzstichtag).

Weitere Schritte werden sich zunächst auf das Einbeziehen zusätzlicher Maßnahmebereiche beschränken. Gelingt es, auf diesem Wege für Evaluationen und/oder andere Forschungszwecke brauchbare Scientific Use Files zu erstellen – was zu erhoffen ist, aber vor der Erprobung im Rahmen dieses Pilotprojektes nicht sichergestellt werden kann – wären weitere Datenanreicherungen (z.B. mit der Beschäftigtenhistorik und zum Verbleib in sozialversicherungspflichtiger Beschäftigung) anzustreben. Im Hinblick auf die Verfügbarkeit dieser Daten und notwendige zeitliche Vorläufe kann die Bildung eines entsprechenden Scientific Use File im Rahmen des Pilotprojektes nur konzeptionell vorbereitet, aber noch nicht realisiert werden – dies müßte nachfolgend im Forschungsdatenzentrum geschehen.

Die stufenweise zu erstellenden Scientific Use Files sollen im übrigen einen neuen, zusätzlichen, zufällig vergebenen Identifikator enthalten, um eine spätere Re-Anonymisierung in einem Forschungsdatenzentrum zu ermöglichen. Im IAB und später im Forschungsdatenzentrum wird eine Art „Übersetzungstabelle“ vorgehalten. Damit wird Nutzern der Scientific Use Files eine Option eröffnet, im Rahmen laufender Projekte auch auf alternative oder ergänzende Forschungsansätze zurückzugreifen.

---

<sup>5</sup> Nicht in allen Maßnahmebereichen enthalten die Ausgangsdatensätze z.B. ein Qualitätsmerkmal (oder mehrere) – kaum vorstellbar eine Kontrollgruppenbildung ohne solch ein Merkmal.

#### 4. Überführung in Forschungsdatenzentrum

Den Vorschlägen des KVI-Gutachtens und des Gründungsausschusses des Rates für Sozial- und Wirtschaftsdaten entsprechend soll in Abstimmung mit externen Wissenschaftlern in der BA ein Forschungsdatenzentrum aufgebaut werden, ein entsprechender Förder-Antrag an das BMBF soll in Kürze gestellt werden.

Soweit im Rahmen des vorliegenden Pilotprojektes Datenerschließungsarbeiten zugunsten externer Nutzer vorgezogen und verstärkt werden, können sie später zumindest z.T. auslaufen. Gleiches gilt hinsichtlich der „Machbarkeitsstudie“ (Erstellung eines Scientific Use Files). Weitere Arbeiten zur Datenerschließung und Unterstützung externer Projekte sollen nach Abschluß des separat finanzierten Pilotprojektes im Finanzierungsrahmen des Datenzentrums weitergeführt werden.<sup>6</sup> Für die Nutzergruppe „externe Wirkungsforschung“, die mit dem vorliegenden Pilotprojekt vor allem angesprochen wird<sup>7</sup>, ist nach gegenwärtigen Überlegungen längerfristig etwa die Hälfte der für das Pilotprojekt veranschlagten Kapazitäten zu rechnen.

---

<sup>6</sup> Organisatorisch wird eine frühestmögliche Verknüpfung angestrebt.

<sup>7</sup> Andere Nutzergruppen eines BA-Forschungsdatenzentrums sind z.B. Biografie-Forscher oder am Betriebspanel des IAB interessierte Institute.