

FDZ-Jahresbericht 2007

Stefan Bender, Benedikt Hartmann, Dagmar Herrlinger

Inhaltsverzeichnis

Vorbemerkung	3
1. Allgemeine Aufgaben	4
2. Basisinformationen	5
3. Das serviceorientierte FDZ	7
4. Das internationale FDZ	12
5. Das forschende FDZ	15

Vorbemerkung¹

Nach den beiden Evaluationen 2006 durch den Rat für Sozial- und Wirtschaftsdaten und 2007 durch den Wissenschaftsrat möchte das Forschungsdatenzentrum der Bundesagentur für Arbeit im Institut für Arbeitsmarkt- und Berufsforschung (FDZ) in Zukunft regelmäßig über seine Aktivitäten berichten. Wir starten mit diesem FDZ-Jahresbericht 2007, der keine Hochglanzbrochure mit vielen Fotos ist, sondern auf wenigen Seiten die wesentlichen Ereignisse der letzten 12 bzw. 24 Monate zusammenfassen soll. Der Jahresbericht dient auch der Transparenz für unsere Nutzer², die die Existenzberechtigung des FDZ begründen. Wir erhoffen uns dadurch Verständnis für unsere Prioritätensetzung und dafür, dass wir nicht alle Wünsche erfüllen konnten. Das Gutachten des Wissenschaftsrats bestätigt uns in vielen Punkten, den bisherigen Kurs fortzuführen (siehe Zitate im Text). Weitere Anregungen für unsere Arbeit sind natürlich erwünscht.

Der Bericht teilt sich in die Kapitel „Allgemeine Aufgaben“, „Basisinformationen“, „Das serviceorientierte FDZ“, „Das internationale FDZ“ und zum guten Schluss „Das forschende FDZ“.

Wir wünschen allen Nutzern, Interessenten und Unterstützern viel Spaß beim Lesen des Jahresberichts 2007.

¹ Wir bedanken uns bei allen Mitarbeiterinnen und Mitarbeiter des FDZ für die hilfreichen Hinweise und natürlich für die tolle Arbeit in 2007.

² Die weiblichen und männlichen Mitarbeiter des FDZ haben das Thema „Gleichbehandlung und Genus“ diskutiert. Es herrschte Einigkeit, dass sich die Gleichbehandlung von Mann und Frau auch in der Sprache widerspiegeln muss. Insbesondere müssen Frauen sprachlich angemessen repräsentiert sein.

Im Folgenden wird die männliche Form verwendet, obgleich uns Produzentinnen, (Gast-)Wissenschaftlerinnen, Nutzerinnen, Mitarbeiterinnen, Kolleginnen und Autorinnen ganz besonders willkommen sind.

1. Allgemeine Aufgaben

Das Forschungsdatenzentrum (FDZ) der BA im IAB bereitet Individualdatensätze auf, die im Bereich der Sozialversicherung und in der Arbeitsmarkt- und Berufsforschung entstehen und stellt sie für wissenschaftliche Zwecke – vorwiegend für externe Forscherinnen und Forscher – zur Verfügung. Das FDZ fördert somit den Einsatz konkurrierender wissenschaftlicher Ansätze und trägt damit auch zur Qualitätssicherung der Daten bei. Dabei versteht sich das FDZ als Mittler zwischen den Datenproduzenten wie der BA, den Forschungsbereichen und Projektgruppen des IAB sowie der externen Wissenschaft.

Das FDZ hat hierzu transparente und standardisierte Zugangsregelungen unter Einhaltung der geltenden Datenschutzbestimmungen entwickelt, führt Datenaufbereitungen, -aktualisierungen und -prüfungen durch, dokumentiert detailliert und umfassend die bereitgestellten Daten und berät individuell über Datenzugang, Datenhandling und Analysemöglichkeiten sowie zur Reichweite und Gültigkeit der Daten.

Zu den Mikrodatensätzen des FDZ gehören (Stand: Februar 2008) das IAB-Betriebspanel, die IAB-Beschäftigtenstichproben (IABS), das BA-Beschäftigten-Panel (BAP), die Stichprobe der Integrierten Erwerbsbiographien des IAB (IEBS), das Betriebs-Historik-Panel (BHP) und der Linked-Employer-Employee-Datensatz des IAB (LIAB). Diese Datensätze sind aufgrund des breiten Merkmalspektrums und der großen Fallzahlen in der Arbeitsmarktforschung von internationaler Bedeutung.

Das bestehende Datenangebot wird laufend aktualisiert und durch neue Datensätze erweitert, bestehende Datenmerkmale werden bereinigt und neue Merkmale hinzugefügt.

Mit seinem Internetauftritt (<http://fdz.iab.de>), den dort verfügbaren Dokumentationen und Arbeitshilfen wie dem FDZ-Datenreport³ oder dem FDZ-Methodenreport⁴ (vgl. auch Anlage 1) sowie mit seinen Workshops und Nutzerkonferenzen erleichtert das FDZ für externe Forscherinnen und Forscher die Arbeit mit den Datensätzen. So richtete das FDZ 2007 gemeinsam mit den Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder den ersten Nachwuchsworkshop des Rats für Sozial- und Wirtschaftsdaten für junge Forscherinnen und Forscher aus. Ziel des Workshops war das Erlernen bzw. Verbessern des praktischen Umgangs mit Mikrodaten für die eigenen Qualifikationsarbeiten. Alle zwei Jahre (letztmalig im Dezember 2006) richtet das FDZ eine Nutzerkonferenz aus, auf der Nutzerinnen und Nutzer die Ergebnisse ihrer Forschungsarbeiten in der BA / im IAB vorstellen.

³ In 2005 wurden neun, in 2006 sechs und in 2007 vier Datenreporte veröffentlicht.

⁴ In 2005 wurden vier, in 2006 fünf und in 2007 neun Methodenreporte veröffentlicht, darunter auch Beiträge von externen Forschern.

2. Basisinformationen

Mit dem Aufbau des FDZ wurde im Dezember 2003 begonnen, da es bis dato an einem systematischen Zugang zu Sozialdaten fehlte. Nach der positiven Evaluation durch den Rat für Sozial- und Wirtschaftsdaten im April 2006 wurde das FDZ auf Dauer als unabhängiges Forschungsdatenzentrum der BA im IAB etabliert. Die Evaluation des Wissenschaftsrats in 2007 bescheinigte dem FDZ, eine international einmalige Einrichtung zu sein:

„Das Forschungsdatenzentrum (FDZ) (Schwerpunkt: Methoden und Datenzugang) ist eine international sichtbare und in Europa einzigartige, unverzichtbare Serviceeinrichtung mit Vorbildcharakter für andere Einrichtungen, die über große, wissenschaftlich bedeutsame Datensätze verfügt.“ (WR-Gutachten, S.55)

Das FDZ ist in seiner Grundfinanzierung mit einem Leiter (AT), fünf wissenschaftlichen Mitarbeiter-, drei nicht-wissenschaftlichen Mitarbeiterstellen und studentischen Hilfskräften (40 Wochenstunden) ausgestattet (vgl. Übersicht 1). Zwei der wissenschaftlichen Stellen sind momentan befristet vergeben. Neben den fünf Stellen für wissenschaftliche Mitarbeiter verfügt das FDZ noch über eine halbe Promotionsstelle als Ermächtigung und eine Graduiertenstelle. Alle Mitarbeiter des FDZ werden momentan dem SGB III zugeordnet.

Durch Drittmittelprojekte (Stand: Februar 2008) sind noch zusätzliche 3,5-Stellen im FDZ besetzt. Weitere 2,5 Drittmittelstellen sind unmittelbar bei Projektpartnern angesiedelt, bei denen das FDZ Mit Antragsteller war (vgl. auch das Kapitel „Das forschende FDZ“).

Übersicht 1: Personalausstattung des FDZ (Stand Februar 2008)

Name, Vorname	Arbeitsschwerpunkte	Finanzierung
Bender, Stefan	Leitung	Grund
Drews, Nils	IABS	Grund
Hartmann, Benedikt	Datenfernverarbeitung, Betreuung der Gastarbeitsplätze, Nutzerdatenbank	Grund
Herrlinger, Dagmar	Veranstaltungen, Internet, Organisation	Grund
Hethy, Tanja	BHP, KombiFiD	Drittmittel
Huber, Martina	IAB-Betriebspanel, WeLL, FAWE-Panel	Drittmittel
Jacobebbinghaus, Peter	LIAB	Grund
Müller, Dana	IAB-Betriebspanel, Metadaten	Grund
Oberschachtsiek, Dirk	IEBS	Grund
Schmucker, Alexandra	IAB-Betriebspanel, BAP, WeLL, Internet	Grund
Scioch, Patrycja	GradAB, KombiFiD	Drittmittel
Seth, Stefan	Datenerstellung, Betreuung der Gastarbeitsplätze	Grund
Spengler, Anja	BHP, KombiFiD	Drittmittel

Das FDZ ist momentan mit 17 Arbeitsplätzen ausgestattet. Durch eine akute Raumnot im IAB, musste in 2007 der zweite Gästeraum aufgelöst werden. Momentan arbeiten somit die Gäste im FDZ in einem Raum, der mit vier Arbeitsplätzen ausgestattet ist. Es gab in 2007 Zeiten, in

denen bis zu vier Gäste im FDZ in diesem Raum gearbeitet haben (vgl. Übersicht 2). Im Schnitt waren die Gastarbeitsplätze im FDZ zu 64% ausgelastet.

Übersicht 2: Belegungsplan des Gästeraums in 2007

Kalender- woche	Datum	Gastrechner 1					Gastrechner 2					Gastrechner 3					Gastrechner 4					
		Mo	Di	Mi	Do	Fr	Mo	Di	Mi	Do	Fr	Mo	Di	Mi	Do	Fr	Mo	Di	Mi	Do	Fr	
1	1.1																					
2	8.1																					
3	15.1																					
4	22.1																					
5	29.1																					
6	5.2																					
7	12.2																					
8	19.2																					
9	26.2																					
10	5.3																					
11	12.3																					
12	19.3																					
13	26.3																					
14	2.4																					
15	9.4																					
16	16.4																					
17	23.4																					
18	30.4																					
19	7.5																					
20	14.5																					
21	21.5																					
22	28.5																					
23	4.6																					
24	11.6																					
25	18.6																					
26	25.6																					
27	2.7																					
28	9.7																					
29	16.7																					
30	23.7																					
31	30.7																					
32	6.8																					
33	13.8																					
34	20.8																					
35	27.8																					
36	3.9																					
37	10.9																					
38	17.9																					
39	24.9																					
40	1.10																					
41	8.10																					
42	15.10																					
43	22.10																					
44	29.10																					
45	5.11																					
46	12.11																					
47	19.11																					
48	26.11																					
49	3.12																					
50	10.12																					
51	17.12																					
52	24.12																					
1	31.12																					

PC nicht belegt
 Gastaufenthalt
 kein Gastaufenthalt möglich wegen Feiertagen, internen Veranstaltungen oder Wartungsarbeiten

Seit Dezember 2007 steht den Nutzern des FDZ ein neuer Server zur Verfügung. Dieser hat vier deutlich schnellere Doppelkern-Prozessoren und einen 32 GB Arbeitsspeicher. Durch das Betriebssystem Windows ist die Oberfläche anwenderfreundlicher als bisher. Der alte Unix-

Server mit 16 GB Arbeitsspeicher wird weiter für einzelne Projekte in der Datenfernverarbeitung verwendet, so dass die Erweiterung einer Verdreifachung unserer Rechenkapazität entspricht.

3. Das serviceorientierte FDZ

In 2007 wurde eine Reihe von Daten aktualisiert, die das FDZ in den letzten Jahren aufbereitet und angeboten hat (vgl. Übersicht 3). Neu in das Datenangebot wurde das Betriebs-Historik-Panel (BHP) aufgenommen. Mit der Bereitstellung der Daten gehen natürlich Updates der dazugehörigen FDZ-Datenreporte einher. In den Datenreporten wurde die Darstellung der Merkmale (Häufigkeiten oder Mittelwerte) bei allen Datensätzen auf Stata-log-Dateien umgestellt, um den Aufwand der Word-Formatierung zu sparen. Die Mitarbeiter des FDZ haben das Datenangebot des FDZ in 2006 und 2007 in vielen Vorträgen im In- und Ausland vorgestellt (vgl. Anhang 2).

Übersicht 3: Entwicklungen des FDZ-Datenangebots und der Dokumentationen in 2007

IAB-Betriebspanel	<ul style="list-style-type: none"> – Bereitstellung der Welle 2006 – Neuentwicklung der Testdaten
IAB-Beschäftigtenstichprobe (IABS)	<ul style="list-style-type: none"> – Update der schwach anonymisierten IABS auf 2004 – Überarbeitete Dokumentation (FDZ-Datenreport 3/2007)
Integrierte Erwerbsbiographien (IEBS)	<ul style="list-style-type: none"> – Bereitstellung des Scientific Use Files IEBS-SUF – Dokumentation des IEBS-SUF (FDZ-Datenreport 1/2007)
Betriebs-Historik-Panel (BHP)	<ul style="list-style-type: none"> – Update des BHP <ul style="list-style-type: none"> ○ Ausweitung des Beobachtungszeitraums auf 1975-2005 ○ Erweiterung um neue Merkmale – Dokumentation des BHP 2 (FDZ-Datenreport 4/2007)
Linked Employer-Employee Daten des IAB (LIAB)	<ul style="list-style-type: none"> – Update des LIAB-Querschnittmodells auf 2005 einschließlich der Erweiterung um neue Merkmale – Überarbeitung der Testdaten des Querschnittmodells – Überarbeitung des LIAB-Datenhandbuchs (FDZ-Datenreport 2/2007)

Die wachsende Bedeutung des FDZ lässt sich mit vielen Zahlen belegen. Hierzu wird eine eigene Datenbank geführt, in der die Aktivitäten des FDZ dokumentiert werden.

Um mit den Daten des FDZ arbeiten zu können, muss vor dem ersten Datenkontakt ein Antrag auf Datennutzung einschließlich Projektbeschreibung abgegeben werden. Nach der Genehmigung durch das Bundesministerium für Arbeit und Soziales wird dann ein Nutzungsvertrag zwischen dem Wissenschaftler und dem IAB abgeschlossen. In Übersicht 4 ist die Anzahl der Genehmigungen nach Datensatz und Datenzugang in 2005-2007 zu sehen. Die Anzahl der

Projekte steigt kontinuierlich von 81 (in 2005) auf 116 (in 2007). Zu beachten ist auch, dass die Projekte i.d.R. länger als ein Jahr dauern und somit Projekte aus 2005 und 2006 auch in 2007 weiter bearbeitet wurden. Die drei am meisten nachgefragten Datensätze sind das IAB-Betriebspanel (mit konstant hohen Zahlen), die IAB-Beschäftigtenstichprobe – wobei hier auch der Gastaufenthalt eine Rolle spielt – und der LIAB (mit einem hohen Anstieg der Verträge über die Jahre).

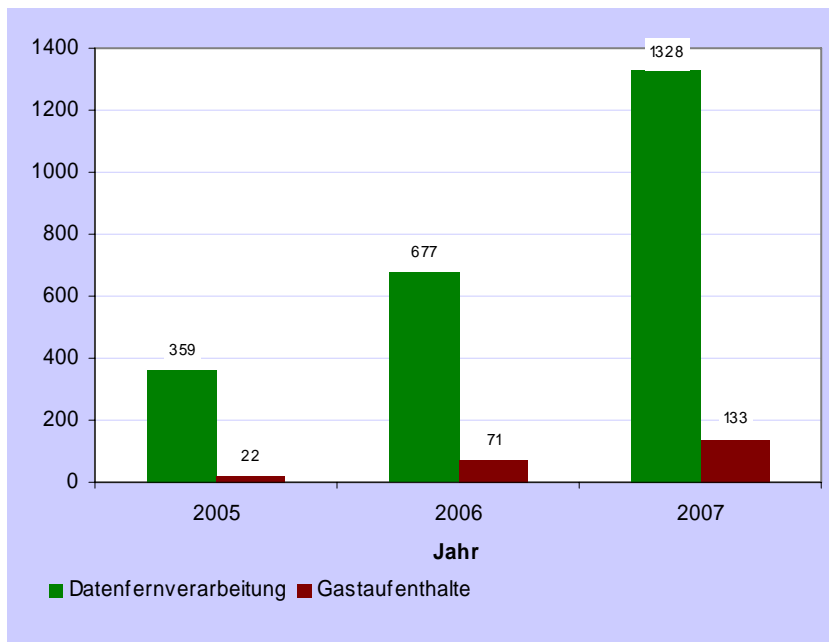
Übersicht 4: Anzahl der Genehmigungen neuer Projekte/Verträge nach Datensatz und Datenzugang

Datensatz	Datenzugang	2005	2006	2007	2005 - 2007*1
keine Zuordnung	Datenfernverarbeitung	0	0	2	2
Beschäftigtenpanel	Datenfernverarbeitung	1	3	9	13
Beschäftigtenpanel	Gastaufenthalt	0	1	1	2
Beschäftigtenpanel	Scientific Use File	8	0	4	12
Betriebspanel	Datenfernverarbeitung	29	36	35	100
Betriebspanel	Gastaufenthalt	2	8	7	17
BHP	Gastaufenthalt	0	2	7	9
IABS	Gastaufenthalt (evtl. anschl. Datenfernverarbeitung)	1	7	8	16
IABS	Scientific Use File	31	26	19	76
IEBS	Gastaufenthalt	0	3	3	6
LIAB	Gastaufenthalt (evtl. anschl. Datenfernverarbeitung)	9	14	21	44
		81	100	116	297

* Summe der bisher in den Jahren 2005 bis 2007 abgeschlossenen Verträgen für den Zugang zu Sozialdaten

Zwei weitere, sehr wichtige Kenngrößen sind die Anzahl der Datenfernverarbeitungen – hier senden Wissenschaftlerinnen und Wissenschaftler Auswertungsprogramme an das FDZ – und die Anzahl der (meist mehrtägigen) Gastaufenthalte im FDZ (siehe Graphik 1). Beide Zahlen haben sich im Vergleich zu 2006 in etwa verdoppelt und zu 2005 sogar vervier- bzw. nahezu verfünffacht.

Graphik 1: Entwicklung der Datenfernverarbeitung und der Gastaufenthalte im FDZ



Diese Zahlen lassen sich für die einzelnen Datensätze ausgeben (vgl. Übersicht 5). Hier zeigt sich, dass der LIAB der am meisten nachgefragte Datensatz des FDZ ist. Über die Hälfte aller Auswertungen in der Datenfernverarbeitung bezieht sich auf ihn, ein Viertel auf das IAB-Betriebspanel. Hervorzuheben ist, dass auch mit dem BHP – obwohl erst in 2007 angeboten – schon 78 Auswertungsläufe via Datenfernverarbeitung in 2007 durchgeführt wurden.

Übersicht 5: Anzahl der Datenfernverarbeitungen nach Datensätzen

	BAP	Betriebs-panel	BHP	IABS	IEBS	LIAB	keine Angabe	Summe/Jahr
2005	1	196			6	155	1	359
2006	13	257		26	11	370		677
2007	55	368	78	79	34	713	1	1328
Summe	69	821	78	105	51	1238	2	

Bei den Gastaufenthalten zeigt sich, dass die steigende Anzahl der Aufenthalte nicht mit einer steigenden Anzahl der Gäste einhergeht (vgl. Übersicht 6). Die Anzahl der Gäste hat sich von 2006 auf 2007 nur leicht erhöht (von 33 auf 39 Gäste). Die durchschnittliche Dauer ist bei drei Tagen geblieben. Es hat sich allerdings die Verteilung der Dauern in 2007 gegenüber den beiden Vorjahren geändert. So hatte das FDZ in 2007 viele Gäste mit sehr kurzen Aufenthalten aus der Universität Erlangen-Nürnberg und einige ausländische Gäste mit längeren Aufenthalten von teilweise über einen Monat.

Übersicht 6: Kenngrößen zu den Gastaufenthalten im FDZ

	Aufenthalte*	Gäste	Durchschnittliche Dauer (Tage)
2005	22	13	3
2006	71	33	3
2007	133	39	3
Summe	226	85	

* Gastaufenthalte eventuell mit anschließender Datenfernverarbeitung

Die Nachfrage ist absolut und auch in ihrer Entwicklung sehr positiv zu bewerten. Daher kommt der Wissenschaftsrat (WR) in seinem Gutachten zum IAB zu folgendem Schluss (WR-Gutachten, S. 50): „Das Forschungsdatenzentrum (FDZ) hat sich in kurzer Zeit zu einer unverzichtbaren und stark nachgefragten Serviceeinrichtung für die einschlägigen deutschen und internationalen scientific communities entwickelt.“

In Übersicht 7 sind die durchschnittlichen Rechenzeiten von Programmen in der Datenfernverarbeitung nach dem jeweiligen Datensatz aufgelistet. Man sieht sehr schön, dass der am meisten nachgefragte LIAB mit ca. 13,7 Stunden die größte durchschnittliche Rechenzeit pro Auswertung benötigt.

Übersicht 7: Durchschnittliche Rechenzeit (min) der Programme in der Datenfernverarbeitung

	BAP	Betriebs-panel	BHP	IABS	IEBS	LIAB
2005	12	155				723
2006	50	67		678	81	1016
2007	29	136	88	685	435	821

Das Warten auf die Ergebnisse wird hauptsächlich durch die Rechenzeit verursacht, da die datenschutzrechtliche Prüfung der Auswertungen vergleichsweise gering ist (vgl. Übersicht 8).

Übersicht 8: Durchschnittliche Prüfzeit (min) der Auswertungen in der Datenfernverarbeitung

	BAP	Betriebs-panel	BHP	IABS	IEBS	LIAB
2005	60	31				21
2006	57	51		21	15	24
2007	32	22	18	21	18	12

Diese kurzen Prüfzeiten führen aber in der Masse zu erstaunlichen Aufwänden an Arbeitszeit wie dies Übersicht 9 verdeutlicht. So wurden für die Datenschutzkontrolle der Auswertungen via Datenfernverarbeitung insgesamt ca. 48 Arbeitstage (bei 8 Stunden pro Tag) benötigt. Natürlich gibt es einige Ausreißer nach oben, die Prüfzeiten von mehreren Tagen benötigen. Hier hat sich die Praxis bewährt, die Nutzer zu kontaktieren und die Menge des Outputs zu beschränken.

Übersicht 9: Summe der Prüfzeiten (min) durch Auswertungen in der Datenfernverarbeitung

	BAP	Betriebs-panel	BHP	IABS	IEBS	LIAB	Summe/Jahr
2005	60	3348				1440	4848
2006	690	13145		535	165	8980	23515
2007	1771	8423	1448	1758	625	8966	22991
Summe	2521	24916	1448	2293	790	19386	

Die Bedeutung des FDZ für die wissenschaftliche Landschaft ist somit hoch. Dies zeigen auch die Publikationen, die mit den Daten des FDZ entstanden sind und die der IAB-Servicebereich „Dokumentation und Bibliothek“ in einer Literaturdatenbank dokumentiert, die im Internet zur Verfügung steht. So wurden bisher⁵:

- mit der IAB-Beschäftigtenstichprobe 246
- mit dem BA-Beschäftigtenpanel 15
- mit dem Betriebs-Historik-Panel 6
- mit den Linked-Employer-Employee-Daten des IAB 82
- mit der Stichprobe der Integrierten Erwerbsbiografien des IAB 50 und
- mit dem IAB-Betriebspanel 1.999

Beiträge von Wissenschaftlern innerhalb und außerhalb des IABs veröffentlicht.

⁵ Die Zahlen beziehen sich auf alle Veröffentlichungen mit dem jeweiligen Datensatz für den Zeitraum ab der ersten Verfügbarkeit des Datensatz im IAB. Einige der Datensätze waren schon weit vor der Entstehung des FDZ im und außerhalb des IAB verfügbar.

