

Forschungsdatenzentrum

der Bundesagentur für Arbeit
im Institut für Arbeitsmarkt-
und Berufsforschung

FDZ

FDZ-Methodenreport

04/2012

DE

Methodische Aspekte zu Arbeitsmarktdaten

Arbeiten und Lernen im Wandel

Teil V: Die Codierung der offenen Angaben zur
beruflichen Tätigkeit, Ausbildung und Branche

Katrin Drasch,
Britta Matthes,
Manuel Munz,
Wiebke Paulus,
Margot-Anna Valentin



Bundesagentur für Arbeit

Arbeiten und Lernen im Wandel

Teil V: Die Codierung der offenen Angaben zur beruflichen Tätigkeit, Ausbildung und Branche

Katrin Drasch (Institut für Arbeitsmarkt- und Berufsforschung Nürnberg)

Britta Matthes (Institut für Arbeitsmarkt- und Berufsforschung Nürnberg)

Manuel Munz (Institut für Arbeitsmarkt- und Berufsforschung Nürnberg)

Wiebke Paulus (Institut für Arbeitsmarkt- und Berufsforschung Nürnberg)

Margot-Anna Valentin (Bundesagentur für Arbeit)

Inhaltsverzeichnis

Zusammenfassung	4
Abstract.....	4
1 Codierung der Angaben zur beruflichen Tätigkeit	5
1.1 Die DKZ der Bundesagentur für Arbeit (BA).....	6
1.2 Entwicklung und Erprobung des IAB Verfahrens zur Berufscodierung	7
1.3 Automatische Codierung.....	8
1.4 Codierung der Berufsangaben ohne automatischen Treffer	9
1.4.1 Manuelle Codierung durch Codierer	10
1.4.2 Codierung schwieriger Fälle.....	11
1.5 Codiererschulung	13
1.6 Qualitätssicherung	14
2 Codierung der Wunschberufe	15
3 Codierung der Angaben zu Ausbildungen.....	16
3.1 Codierung der Ausbildungen ohne Kurse	16
3.2 Codierung der Kurse	17
4 Codierung der Branchen	18
4.1 Die Klassifikation der Wirtschaftszweige nach der WZ08	18
4.2 Die Codierung der Branchenangaben.....	19
4.2.1 Manuelle Codierung ohne Hilfsangaben	19
4.2.2 Manuelle Codierung mit Hilfsangaben	21
4.2.3 Zuordnungsentscheidungen	22
5 Fazit.....	24
Literaturverzeichnis	25
Anhang	26

Zusammenfassung

Die Studie Arbeiten und Lernen im Wandel (ALWA) wurde von September 2007 bis März 2008 im Auftrag des IAB durchgeführt. Die Zielsetzung dieser repräsentativen, telefonischen Bevölkerungsumfrage bestand darin, die Bildungs- und Erwerbsverläufe von 10.000 Erwachsenen zwischen 18 und 50 Jahren systematisch retrospektiv zu erfassen (vgl. Kleinert et al. 2011, Antoni et al. 2010). In diesem Zusammenhang wurden auch Informationen über die ausgeübte Tätigkeit, Ausbildung und die dazugehörige Branche gesammelt. Ziel der Codierung war es, die offen erfassten Angaben aus ALWA zur beruflichen Tätigkeit, zum Wunsch- und Ausbildungsberuf und die jeweilige Branchenangabe zu codieren. In diesem Bericht wird zuerst die theoretische Grundlage für die Codierung der Angaben zur beruflichen Tätigkeit mit Hilfe Dokumentationskennziffer (DKZ) der Bundesagentur für Arbeit (BA) erläutert. Anschließend wird das Verfahren der technischen Umsetzung der Codierung dargestellt, das sich in mehrere Schritte gliedert: einer automatischen Codierung, einer manuellen Codierung und einer Bearbeitung problematischer Fälle. Danach wird das an die Berufscodierung angelehnte Verfahren für Wunsch- und Ausbildungsberufe erklärt. Außerdem wird das Verfahren zur Codierung der Branchen mit der Klassifikation der Wirtschaftszweige, Ausgabe 2008 (WZ08) vorgestellt. Abschließend wird kurz auf die Schulung der Codierer und Qualitätssicherung der Codierung eingegangen und ein Fazit gezogen.

Abstract

The IAB-ALWA study (Work and Learning in a Changing World) was conducted between September 2007 and March 2008 on behalf of the IAB. The aim of this representative telephone survey was to collect systematically educational and employment histories of 10,000 adults who were between 18 and 50 years old. The aim of the coding process was to generate codes from the open text answers of questions about occupations, occupational ambition, vocational training and economic sector. First the theoretical basis of the coding process of the open text answer of the occupation with the help of the Dokumentationskennziffer (documentation figure (DKZ) from the Federal Employment Agency will be outlined. Next the technical implementation is described which can be organized into several steps: an automatic coding, a manual coding and the handling of problematic cases. Afterwards the coding process for vocational training and courses is described. Additionally, the coding process for economic sectors with the help of the German classification WZ08 (Klassifikation der Wirtschaftszweige, Ausgabe 2008) is illustrated. Finally, the coder training and the quality assurance of the coding is explained and a conclusion is drawn.

Keywords: ALWA, Codierung beruflicher Angaben und Branche, Coding of occupation and branch

An dieser Stelle möchten wir uns bei den zahlreichen Mitarbeiterinnen und Mitarbeitern bedanken, die das Gelingen des Projektes sichergestellt haben: Katharina Derr, Melanie Götz, Kerstin Hecker, Stefanie Jung, Anna Kronhöfer, Alexander Kuppinger, Manuel Munz, Michael Schottenhammer, Claudia Swinnes und Thomas Ullherr.

1 Codierung der Angaben zur beruflichen Tätigkeit

In bisherigen Studien wurden die Angaben zur beruflichen Tätigkeit zumeist vom Zentrum für Umfragen und Methoden (ZUMA) codiert. Die ZUMA greift bei der Codierung der Angaben zur beruflichen Tätigkeit auf ein Verfahren zurück, das sowohl auf computerunterstützte Codierung als auch auf Codierung durch geschulte Fachkräfte basiert. Durch die computerunterstützte Codierung der ZUMA konnten in der Vergangenheit ca. 50 % der Fälle automatisch codiert werden, ca. 50 % der Fälle mussten manuell nachbearbeitet werden.¹

Alternativen zur Codierung durch die ZUMA gibt es nur wenige. Für das Sozio-Ökonomische Panel (SOEP) wurden beispielsweise die Angaben zur beruflichen Tätigkeit von infratest Sozialforschung für die Panelwellen 1984 bis 2001 neu codiert, um eine bessere Vergleichbarkeit zwischen den Panelwellen herzustellen. Infratest greift dabei sowohl auf eine alphabetische Gliederung der „Klassifizierung der Berufe“ in der Ausgabe von 1992 (KIdB 1992) zurück, als auch auf eine Datei, die vom IAB zur Verfügung gestellt wurde (vgl. Hartmann/Schütz 2002). Infratest hat dieses Verfahren in den letzten Jahren weiter verbessert und es auch im Berichtssystem Weiterbildung (BSW) sowie in der BIBB/BAuA-Erwerbstätigenbefragung 2005/2006 eingesetzt.

Die hier vorgestellte Vorgehensweise des IAB zur Codierung der beruflichen Tätigkeit stellt eine Alternative zum Verfahren der ZUMA und Infratest dar. Ziel dieses Vorgehens ist es, die Angaben zur beruflichen Tätigkeit in die Dokumentationskennziffer (DKZ) der Bundesagentur für Arbeit (BA) zu überführen. Als Datenbasis für die automatische Codierung wurde dabei die Datenbank der Auflistung aller in der BA bekannten aktuellen und veralteten Berufe verwendet. Das Vorgehen des IAB bietet somit einen entscheidenden Vorteil gegenüber den anderen Verfahren. Es greift auf eine ständig gepflegte und auch in der Praxis verwendete Datenbank zurück, die nicht explizit zum Zweck der Codierung von Angaben zur beruflichen Tätigkeit erstellt wurde.

Ziel der Codierung der Tätigkeitsangaben war die möglichst exakte und detaillierte Verschlüsselung der Berufsangaben als DKZ. Für diese DKZ sind Umsteigercodes in die bekannten Berufsklassifikationen der nationalen „Klassifikation der Berufe“ des Statistischen Bundesamtes in der Ausgabe von 1988 (KIdB 1988) und einer internationalen Klassifikation der Berufsgattungen, der „International Standard Classification of Occupations (ISCO)“ des „International Labour Office (ILO)“ aus

¹ Für eine detaillierte Erläuterung der Berufsvercodung, wie sie von der ZUMA durchgeführt wird, siehe das Handbuch für die Berufsvercodung (vgl. Geis & Gantz 2007).

dem Jahr 1988 (ISCO 1988), vorhanden. Mittels dieser Berufsklassifikationen und anderer Informationen aus der Befragung, wie der beruflichen Stellung, ist beispielsweise eine Generierung der Prestigewerte nach Treiman (SIOPS), Ganzeboom (ISEI) und Wegener (MPS) möglich (Geis 2009).² Das ermöglicht wiederum die Untersuchung von Forschungsfragen zu beruflicher Mobilität und berufliche Segregation. Im Folgenden wird kurz die Arbeitsgrundlage für die Codierung, die DKZ der Bundesagentur für Arbeit (BA) dargestellt, bevor auf das IAB Verfahren zur Berufscodierung und dessen Ergebnisse eingegangen wird.

1.1 Die DKZ der Bundesagentur für Arbeit (BA)

Die DKZ der BA wird im Rahmen der Vermittlung und Beratung der BA unter anderem zur Verschlüsselung von Berufsangaben bei Stellen- und Bewerberangeboten in der JOBBÖRSE, dem Online Stellenportal der BA, verwendet und stellt die Grundlage für das Matching von Berufen und Stellen dar. Sie entstammt also dem operativen Geschäft der BA und wird daher ständig aktualisiert. Zudem wird die Kennziffer auch in der Statistik der BA verwendet, um Ausbildungs- und Tätigkeitsbezeichnungen zu verschlüsseln (vgl. Paulus et al. 2010).

Die DKZ-Datenbank beinhaltet eine umfassende Sammlung aller in Deutschland verwendeten Berufs- und Ausbildungsbezeichnungen, deren DKZ und weiterführenden Informationen. Die permanent aktualisierte Datenbank besteht aus Metadaten vieler Informationssysteme der BA (BERUFENET, KURSNET, VerBIS u.a.). Alle verwendeten Ausbildungs- und Tätigkeitsbezeichnungen wurden durch die Bundesagentur für Arbeit (BA) definiert, außerdem wurden diesen Bezeichnungen verschiedene Suchbegriffe zugeordnet. Die Informationen beinhalten neben der DKZ auch die genauen Berufsbezeichnungen sowie die berufskundliche Gruppe in der Systematik der BA. Die berufskundliche Gruppe zeigt an, welchem Ausbildungsniveau die Tätigkeit entspricht. Außerdem ist es möglich, zwischen Tätigkeiten und Ausbildungen zu differenzieren. Die Datenbank enthält neben aktuellen Bezeichnungen auch nicht mehr gebräuchliche Bezeichnungen von Berufen. Die Datenbank ist für die Öffentlichkeit im Internet unter <http://download-portal.arbeitsagentur.de/files/> zugänglich.³ Derzeit enthält diese Datenbank ca. 11.000 Einträge auf der Ebene der Klassifizierung der Berufe.

² Für einen Überblick über national und international gebräuchliche Skalen wird an dieser Stelle auf Hoffmeyer-Zlotnik und Geis (2003) verwiesen.

³ Inzwischen ist die DKZ 8-stellig, weil die ersten 5 Stellen der neu entwickelten KIdB 2010 (Paulus et al. 2010) entsprechen. Weitere Informationen zur DKZ sind bei den Autoren erhältlich.

Die komplette DKZ besteht aus sieben Ziffern. Die ersten vier Stellen sind mit dem Viersteller der KldB 1988 identisch. Die erste der drei weiteren Ziffern ermöglicht es, Tätigkeiten (Ziffern 1 und 2) von Ausbildungen (Ziffern 8 und 9) zu unterscheiden. Die beiden anderen Ziffern werden vergeben, ohne damit eine Systematisierung zu beabsichtigen. Sie dienen lediglich einer systemfreien, weiteren Differenzierung der Berufsbezeichnungen.

1.2 Entwicklung und Erprobung des IAB Verfahrens zur Berufscodierung

Die Angaben zur beruflichen Tätigkeiten wurden in ALWA mittels zweier offenen Fragen erhoben, wobei die zweite Nennung die erste näher spezifizieren sollte. Die Fragen, die gestellt wurden, lauteten wie folgt: 1. „Sagen Sie mir bitte, welche berufliche Tätigkeit Sie da ausgeübt haben!“ und 2. „Können Sie mir diese berufliche Tätigkeit noch konkreter benennen? Hat das, was Sie gemacht haben, noch eine genauere Bezeichnung?“. Diese Fragen wurden nicht nur für derzeitige berufliche Tätigkeiten gestellt, sondern auch für alle in der Vergangenheit ausgeübten beruflichen Tätigkeiten und alle Nebentätigkeiten.

Ziel der Berufscodierung war es, für jede Erwerbsepisode einen Code zu erhalten. Für die Zuordnung dieser offenen Angaben zu der in der Bundesagentur für Arbeit verwendeten Dokumentationskennziffer (DKZ) wurde ein mehrstufiges Verfahren entwickelt. In einem ersten Schritt erfolgte eine automatische Codierung der Angaben zur beruflichen Tätigkeit, wobei in Schritt 1a nur eindeutige Zuordnungen von beruflichen Tätigkeiten zu Codes (1:1-Verknüpfung, sogenannte Endpunkte) einbezogen wurden und in Schritt 1b die Zuordnung auf alle Suchworte (auch 1:n-Verknüpfung) ausgedehnt wurde. Im zweiten Schritt wurden die nicht automatisch codierbaren Angaben von entsprechend geschulten Codierern manuell codiert. Im dritten Schritt wurden die bis zu diesem Zeitpunkt nicht codierbaren Angaben von Supervisoren geprüft und wenn möglich codiert. In den folgenden Abschnitten werden diese Schritte näher erläutert. Tabelle 1 zeigt die Ergebnisse der Codierung der Angaben zur beruflichen Tätigkeit in den jeweiligen Schritten. Die erste Spalte der Tabelle berichtet die Ergebnisse bezogen auf alle Erwerbsepisoden, die zweite Spalte bezieht sich sowohl auf die erste als auch auf die zweite Angabe zur beruflichen Tätigkeit.

Von den 33.336 Erwerbsepisoden konnten durch die automatische Codierung über die Endpunkte über 11.000 Erwerbsepisoden einem eindeutigen Code zugewiesen werden. Dahinter steht, dass bei 14.000 Angaben zur beruflichen Tätigkeit entweder nur eine Angabe vorhanden war und dieser einem eindeutigen Code zugeordnet werden kann, oder die erste und zweite Tätigkeitsbezeichnung zum gleichen Treffer geführt hat. Für 1.700 Erwerbsepisoden waren Zuordnungsentscheidungen

zu treffen, wobei davon bei 459 Erwerbsepisoden der Treffer zwischen der ersten und zweiten Angabe nicht identisch war. Durch die automatische Codierung über die Suchworte konnten die beruflichen Tätigkeitsangaben von über 8.000 Erwerbsepisoden codiert werden, fast 3.000 direkt und etwa 5.300 durch Auswahl eines der Treffer.

Tabelle 1: Ergebnis der Codierung der Angaben zur beruflichen Tätigkeit

	Erwerbsepisoden insgesamt	Erste und zweite Angabe zu den Erwerbsepisoden
Angaben zur beruflichen Tätigkeit	N=33.336 100 %	N=55.588 100 %
Schritt 1a: Automatische Codierung, Endpunkte	N=12.889 39 %	N=14.635 26 %
Eindeutig zuordenbar	N=11.174 34 %	N= 14.176 25 %
Entscheidung notwendig	N=1.715 5 %	N=459 1 %
Schritt 1b: Automatische Codierung, Suchworte		N=8.316 15 %
Eindeutig zuordenbar		N=2.999 5 %
Entscheidung notwendig		N=5.317 10 %
Schritt 2: Codierung der Angaben ohne Treffer	N=20.477 61 %	
Eindeutig zuordenbar	N=17.442 52 %	
Schritt 3: Entscheidung notwendig oder Problemfall	N=3.005 9 %	

Nach den automatischen Codierungen verblieben etwa 20.000 Erwerbsepisoden, die manuell codiert werden mussten. Da ab diesem Schritt beide Angaben zur Beschreibung der beruflichen Tätigkeit gleichzeitig zur Codierung benutzt wurden, bleibt die zweite Spalte ab hier leer. Durch die manuelle Codierung wurden mehr als 17.000 Erwerbsepisoden codiert. Die übrigen 3.000 wurden durch die Supervision gesichtet und wenn möglich codiert, wenn nicht möglich als nicht codierbar gekennzeichnet.

1.3 Automatische Codierung

Grundlage der Zuordnung der Codes nach der DKZ war die in elektronischen Stichwortregistern verfügbare Auflistung aller in der BA bekannten Berufe. Dies schließt nicht nur aktuelle, sondern auch veraltete Berufsbezeichnungen mit ein. Die offen erhobenen Angaben zur beruflichen Tätigkeit wurden in einem ersten Schritt mit zwei Datenbanken, die jeweils die Berufsangaben enthalten, in folgender Reihenfolge verglichen.

1. Datenbank mit den Bezeichnungen der Endpunkte (auf Siebenstellerebene) in der DKZ bzw. der Zentralen Berufedatei (ZBD): Hier gibt es je-

weils eine männliche und weibliche Kurz- und Langform. Bei ca. 10.500 Endpunkten (1:1-Verknüpfung, d.h. eindeutigen Zuordnungen der beruflichen Tätigkeiten), entstand somit eine Datenbank von ca. 42.000 Einträgen. Dabei wurden sowohl die gültigen als auch die ungültigen (d.h. veralteten) Endpunkte berücksichtigt. (Schritt 1a)

2. Datenbank mit den Suchworten in der DKZ bzw. ZBD: Bei den Suchworten gibt es männliche, weibliche, aber auch neutrale Suchworte. Genutzt wurden nur die männlichen und weiblichen Suchworte. Dies sind jeweils etwa 50.000, so dass eine Datenbank mit ca. 101.000 Einträgen zur Verfügung stand. Den Suchworten sind jeweils mehrere Endpunkte zugeordnet (1:n-Verknüpfung). Für alle Angaben, die mittels der ersten Datenbank zu keinem Treffer führten, wurde ein Abgleich mit der Datenbank der Suchworte vorgenommen. (Schritt 1b)

Das so erstellte Berufsregister wurde anschließend zur automatischen Codierung verwendet. Die erhobenen Angaben zur beruflichen Tätigkeit wurden mit den Einträgen des Registers verglichen. Bei entsprechenden Treffern wurden dann ein DKZ-Siebensteller zugeordnet (Schritt 1a). Beim Abgleich mit den Suchworten wurden entsprechend mehrere DKZ-Siebensteller zugeordnet (Schritt 1b).

Die zwei Angaben zur beruflichen Tätigkeit wurden dabei getrennt voneinander codiert, so dass es auch möglich war, dass die Codierung der ersten Angabe zu einem anderen Ergebnis geführt hat als die Codierung der zweiten Angabe. Bei Berufsangaben mit nur einem einzigen Treffer in den beiden Angaben bzw. mit zwei übereinstimmenden Treffern wurde der jeweilige Treffer bzw. der erste Treffer übernommen. Bei Berufsangaben mit zwei unterschiedlichen Treffern in den beiden Angaben wurde in einem extra Schritt manuell von einem Supervisor eine Entscheidung getroffen. Ebenso wurde bei mehreren DKZ-Siebenstellern, die sich durch die Suchwortabfrage ergaben, manuell eine Entscheidung getroffen.

1.4 Codierung der Berufsangaben ohne automatischen Treffer

Berufsangaben, die nicht automatisch codiert werden konnten, wurden von geschulten Codierern in Codes transferiert. Angaben zur beruflichen Tätigkeit, die in diesem Schritt nicht codiert werden konnten, da entweder keine DKZ oder mehrere DKZs zugeordnet werden konnten, oder sich der Codierer bei der Vergabe des Codes unsicher war, wurden als schwierige Fälle klassifiziert und in einem weiteren Schritt von Supervisoren codiert. Das detaillierte Vorgehen wird in den nächsten beiden Abschnitten erläutert.

1.4.1 Manuelle Codierung durch Codierer

Die Berufsangaben ohne automatischen Treffer wurden manuell von geschulten Codierern codiert (Details zur Codiererschulung sind in Abschnitt 1.5 zu finden). Die Codierer erhielten neben den zwei Berufsangaben aus dem ALWA-Datensatz auch entsprechende Angaben zur differenzierten beruflichen Stellung und zur Branche, die sie bei Bedarf für die Codierung heranziehen konnten. Zudem konnten die Codierer anhand eines Zuordnungsschemas der differenzierten beruflichen Stellung zur berufskundlichen Gruppe eines jeden DKZ-Siebenstellers herausfinden, welche Ausbildung üblicherweise für die Ausübung der entsprechenden beruflichen Tätigkeit benötigt wird.

Folgende Grundregel für die Codierer wurde formuliert:

1. Nach Eingabe der Berufsbezeichnung in die DKZ Datenbank erfolgt ein Abgleich mit der beruflichen Stellung und der berufskundlichen Gruppe. Stimmt die Berufsbezeichnung mit diesen Angaben überein wird der Beruf codiert.

Offensichtliche Tipp- oder Rechtschreibfehler durften von den Codierern übergangen werden. Die Angaben mussten orthographisch korrekt in die Suchmaschine eingegeben werden. Außerdem musste die berufskundliche Gruppe, die der Tätigkeit zugeordnet ist, mit der beruflichen Stellung, die die Zielperson angegeben hat, abgeglichen werden. Falls diese nicht übereinstimmt, wurde nach einer Berufsbezeichnung in entsprechenden berufskundlichen Gruppe gesucht.

Folgende weitere Regeln wurden aufgestellt:

2. Werden mehrere Berufsbezeichnungen angegeben, die aber zu einer oder einer ähnlichen Berufsgruppe gehören, wird die speziellere Tätigkeit codiert. Für den Fall, dass zwei ähnliche Berufsbezeichnungen erfasst wurden (z.B. Lehrer und Gymnasiallehrer), wurde die speziellere Berufsbezeichnung codiert.
3. Für den Fall, dass zwei verschiedene Berufsbezeichnungen in den beiden Angaben erfasst wurden, wurde folgende Regelung formuliert: Wenn zwei DKZ-Siebensteller in Frage kommen und kein Hinweis für die Entscheidung vorliegt oder die Angabe aus einer unspezifischen Bezeichnung besteht, wird die Branche als Entscheidungskriterium hinzugezogen. Falls danach ebenfalls kein Hinweis auf eine Entscheidung vorliegt, werden beide DKZ-Siebensteller an die Supervisoren übergeben, die solche Fälle gesondert betrachten.

Bestand die Angabe zur beruflichen Tätigkeit beispielsweise aus der unspezifischen Bezeichnung „Kaufmann“ waren die Codierer angehalten, die Brancheninformation zu nutzen z.B. „Kaufmann“ (Berufsangabe) und „Industrie“ (Branchenangabe) wurde als „Industriekaufmann“ codiert.

Kann eine der angegebenen Tätigkeiten als hauptsächlich durchzuführende Aufgabe identifiziert werden, so wird diese ebenfalls codiert. Wird als hauptsächlich durchzuführende Aufgabe beispielsweise Maurer identifiziert, so wird nur diese Angabe codiert.

Wenn keine Klärung möglich ist, werden diese Fälle gesondert an die Supervisoren weitergegeben.

Für den seltenen Fall, dass die Angaben zwei sehr unterschiedliche Berufsbezeichnungen erfasst haben (z.B. Schreiner und Koch), wurde die erste der beiden Nennungen codiert (vgl. Geis, 2009, S.7).

1.4.2 Codierung schwieriger Fälle

Im Verlauf des Codierprozesses wurde es notwendig, die oben aufgestellten Regeln zu adaptieren bzw. zu ergänzen. Um eine hohe Konsistenz auch bei der Codierung der schwierigen Fälle zu erreichen, wurden daher weitere spezifische Regeln festgelegt. Als Grundregeln galten weiterhin die Regeln 1 bis 3, die in diesem Schritt noch einmal geprüft wurden. Dennoch war in vielen Fällen eine Entscheidung nicht eindeutig und es mussten weitere Informationen wie der Ausbildungs- oder Erwerbsverlauf hinzugezogen werden.

Regel 3 wurde wie folgt adaptiert:

3. Wenn zwei DKZ-Siebensteller in Frage kommen und kein Hinweis für die Entscheidung vorliegt, werden diese Fälle gesondert betrachtet. Hierbei werden weitere Angaben wie der Ausbildungsberuf, der Erwerbsverlauf oder die Branche hinzugezogen, auf deren Grundlage eine Entscheidung getroffen wird.

Folgende ergänzende Regeln wurden festgelegt:

4. Bei Bezeichnungen, die zu unspezifisch sind, werden weitere Informationen wie der Ausbildungsberuf, der Erwerbsverlauf oder die Branche hinzugezogen und auf deren Grundlage eine Entscheidung getroffen.
5. Werden mehrerer Berufsbezeichnungen angegeben, die aber nicht zur gleichen oder ähnlichen Berufsgruppe gehören, so werden diese Fälle unter Berücksichtigung weiterer Informationen wie Ausbildung und Erwerbs-

verlauf gesondert betrachtet. Kann eine der angegebenen Tätigkeiten als hauptsächlich durchzuführende Aufgabe identifiziert werden, so wird diese codiert.

6. Wird eine Tätigkeit durch mehrere Worte beschrieben, die Tätigkeit wird aber nicht explizit genannt, so wird, wenn eine Ausbildung vorhanden ist, diese als Tätigkeit codiert. Ist keine Ausbildung vorhanden, wird die Tätigkeit als Hilfstätigkeit codiert.
7. Wenn eine Tätigkeit durch Abkürzungen, Bindestriche oder Rechtschreibfehler gekennzeichnet ist, wird die Tätigkeit, sofern sie lesbar ist und sofern eine Ausbildung vorhanden ist, codiert. Sind die Angaben zu unleserlich, wird der Fall in der Einzelfallcodierung gesondert betrachtet.
8. Wird eine berufliche Stellung überbewertet und es erfolgte keine entsprechende Weiterbildung, so wird die Tätigkeit als qualifizierte Tätigkeit codiert.

Der Fall, dass eine berufliche Stellung überbewertet wurde, kam eher selten vor und beschränkte sich vorrangig auf die Berufsbezeichnung „Mechatroniker“. In diesen Fällen lag keine Weiterbildung als Meister vor, daher wurde die Berufsbezeichnung, wie oben beschrieben als Mechatroniker mit qualifizierter Tätigkeit codiert.

9. Wird dagegen eine wissenschaftliche Tätigkeit (z.B. Arzt) als einfache Tätigkeit bewertet, wird sofern ein abgeschlossenes Studium vorhanden ist, die Tätigkeit als wissenschaftliche Tätigkeit bewertet.
10. Wird eine Tätigkeit angegeben, zu der keine Ausbildung absolviert wurde, wird der qualifizierte Beruf codiert, wenn eine langjährige Berufserfahrung vorliegt. Wenn dies nicht der Fall ist, wird die Tätigkeit als Helfertätigkeit codiert.
11. Wird als Tätigkeit Geschäftsführer, Team- oder Gruppenleiter genannt, wird der Fall gesondert betrachtet. Hier wird die Regel aufgestellt, dass die Bezeichnung Geschäftsführer, Team,- oder Gruppenleiter codiert wird, wenn die Person mehr als 10 Mitarbeiter zu führen hat. Ist dies nicht der Fall wird der Ausbildungsberuf oder Studienabschluss codiert.

Dennoch trat mehrmals der Fall auf, dass Tätigkeiten nicht codiert werden konnten, da Angaben unleserlich waren, keinen Sinn ergaben oder da es nicht möglich war, Angaben einer Tätigkeit zuzuordnen. In diesen Fällen galt die Regel:

12. Kann eine Angaben nicht zugeordnet werden, weil sie unleserlich ist oder keinen Sinn ergibt oder die Angabe auch unter Zuhilfenahme aller Zusatz-

angaben und Regeln keiner Tätigkeit zugeordnet werden kann, so wird diese als Missing -4 „Berufsangabe zu ungenau“ gekennzeichnet.

Ein weiteres Problem war die Tatsache, dass nicht zu allen Berufen DKZ-Siebensteller vorhanden sind. Dies ist der Fall, wenn Berufe von der Agentur für Arbeit nicht vermittelt werden, insbesondere weil Personen in diese Ämter, unabhängig von Ausbildung und Erwerbstätigkeit, gewählt werden.

13. Es wurden daher für folgende Tätigkeiten neue Codes vergeben:

Tätigkeiten, in die Personen, unabhängig von Ausbildung und Erwerbstätigkeit, gewählt werden

- Bürgermeister, Stadt-/Landrat: 7613-001
- Funktionär: 7613-002
- Betriebs-/Personalrat: 7632-000
- Ausbildungsleiter für Zivilverteidigung: 8779-000

1.5 Codiererschulung

Die Codierer, die für die Codierung der beruflichen Tätigkeiten und Ausbildungen eingesetzt wurden, waren Studenten bzw. Absolventen von Studiengängen, die einen engen Bezug zur empirischen Sozialforschung haben (z.B. Soziologie bzw. Sozialwissenschaften u.ä.). Die Codierer erhielten eine eintägige Schulung, in der die theoretischen Grundlagen der Berufsklassifikationen und die praktische Herangehensweise an die Codierung erläutert wurden.

Im theoretischen Teil erhielten die Codierer eine Einführung in die DKZ-Datenbank der BA und in die Klassifikation der Berufe 1988 (KIdB), um ihnen das Ziel der Codierung zu erläutern. In einem weiteren Schritt wurden die Arbeitsmittel, die DKZ Information und das Berufenet der BA vorgestellt und die Grundregeln der Codierung erläutert. Abschließend wurde das konkrete Vorgehen anhand mehrerer Beispiele erklärt. Gesondert geschult wurde dabei die Codierung der sog. „Helferberufe“, d.h. Tätigkeiten, die von An- bzw. Ungelernten ausgeführt werden.

Im praktischen Teil der Schulung erhielten die Codierer Beispielfälle, die sie selbstständig in einem begrenzten Zeitraum bearbeiten mussten. Die Codierung dieser Beispielfälle wurde im Gremium besprochen und war zugleich Bestandteil der Qualitätssicherung. Im laufenden Codierprozess wurden auftretende Neuerungen bzw. Unklarheiten unter allen Codierern persönlich bzw. per E-Mail kommuniziert.

1.6 Qualitätssicherung

Zur Qualitätssicherung der codierten Berufsangaben wurden mehrere Maßnahmen durchgeführt. Erstens wurden im zweiten Schritt, der manuellen Codierung durch die Codierer, die ersten 300 Angaben, die ein Codierer codiert hat, von Supervisoren auf Beachtung der Grundregeln geprüft und den Codierern Feedback gegeben. Diese ersten 300 Angaben wurden dann komplett verworfen und am Ende der Codierarbeiten noch einmal von einem anderen Codierer bearbeitet.

Zweitens wurden Zusatzregeln, die im Verlauf der Codierarbeiten entstanden sind, an alle Codierer weitergegeben und schriftlich fixiert. Die von den Zusatzregeln betroffenen Fälle wurden dann erneut überprüft.

Drittens wurden Problemfälle in einer Jour-fixé besprochen und anschließend im Vier-Augen-Prinzip Entscheidungen getroffen. Diese Ergebnisse wurden ebenfalls schriftlich fixiert.

Viertens wurden 20 % der nicht automatisch codierbaren Angaben zur Analyse der Inter-Coder-Reliabilität (blind) doppelt codiert, d.h. es wurden per Zufallsmechanismus offene Angaben ausgewählt, die ohne Kenntnis der bisherigen Codierung von den Codierern erneut codiert wurden. Anhand dieser Maßzahlen wurden Inter-Coder-Reliabilitäten bestimmt. Diese Reliabilität bezeichnet also den Grad der Übereinstimmung der Ergebnisse zweier Codierer, die die Angaben unabhängig voneinander codiert haben.

Tabelle 2: Ergebnisse der Inter-Coder-Reliabilität

Reliabilität	Codierung der offenen Angaben zur beruflichen Tätigkeit
KIdB 2010 - 7-Steller Ebene	50 %
KIdB 2010 - 4-Steller Ebene	65 %
KIdB 2010 - 2-Steller Ebene	79 %
ISCO 1988 - 4-.Steller Ebene	70 %

Die Inter-Coder-Reliabilitäten wurde auf verschiedenen Ebenen berechnet. Auf der am meisten differenzierten Ebene, dem DKZ-Siebensteller beträgt diese Reliabilität knapp über 50 %. Auf 4-Steller Ebene beträgt die Reliabilität bereits 65 % und auf der Ebene der 2-Steller 79 %. Verwendet man die ISCO Klassifikation auf Vierstellerebene aus dem Jahr 1988, ergibt sich eine Reliabilität von 70 %.

Diese Werte erscheinen auf den ersten Blick sehr niedrig. Jedoch beziehen sie sich nur auf die nicht automatisch codierbaren Angaben, d.h. insgesamt wird der Fehler deutlich kleiner sein, da die automatisch codierten Angaben eine geringere Fehlerwahrscheinlichkeit haben. Allerdings zeigt sich hier auch, mit welchen Problemen die Klassifizierungen allgemein zu kämpfen haben: Oftmals ist eine trenn-

scharfe Zuordnung auch unter Zuhilfenahme sämtlicher zur Verfügung stehenden Informationen nicht möglich. Insbesondere betrifft dies neue Berufsfelder, vor allem Dienstleistungsberufe. An einem Beispiel illustriert, bedeutet dies folgendes: Die Angabe Call-Center Telefonistin lässt sich sowohl als Call-Center-Agent/in (7032-101) aber auch als Telefonist/in (7341-100) codieren. Beide Codierungsmöglichkeiten sind richtig, wenn man nicht im Detail klären bzw. nachfragen kann, was eine Person genau macht. Allerdings ergibt sich auf keiner der Ebenen eine Übereinstimmung, falls der Erstcodierer nach der ersten Möglichkeit, der Zweitcodierer nach der zweiten Möglichkeit codiert hat. Insgesamt wird deutlich, dass eine neue Klassifikation der Berufe, wie sie derzeit von der Bundesagentur für Arbeit (vgl. Paulus et al. 2010) erarbeitet wird, unumgänglich ist, um trennscharfe Ergebnisse zu erreichen.

2 Codierung der Wunschberufe

Im Anschluss an die Befragung zur Schulgeschichte in ALWA wurde erfasst, welchen Beruf die Zielpersonen zum Zeitpunkt ihres Schulabschlusses einmal ausüben wollten (Wunschberuf). Diese Angaben wurden ebenfalls mit Hilfe der DKZ analog dem Verfahren zur Codierung der ausgeübten Tätigkeiten codiert.

Tabelle 3: Ergebnis der Codierung der Angaben zum Wunschberuf

	N	Anteil
Angaben zum Wunschberuf	8.980	100 %
Automatische Codierung	7.993	89 %
Codierung der Angaben ohne Treffer	987	11 %
Eindeutig zuordenbar	789	9 %
Problemfall	198	2 %

Es sollte nur ein Hauptwunschberuf erfasst werden und bei 8.980 Personen von 10.177 Befragten, die diese Frage beantworten sollten, war eine Textangabe vorhanden. Davon konnten 7.993 (89 %) automatisch codiert werden. Diese Anzahl ist deswegen so hoch, da im Vergleich zu ausgeübten Tätigkeiten relativ unspezifische Angaben gemacht wurden, die leicht zu codieren waren. Nach der automatischen Codierung erfolgte eine manuelle Codierung. Da pro Angabe nur ein exakter Treffer möglich war, vereinfachte sich das Verfahren entsprechend. Analog zur Codierung der beruflichen Tätigkeiten erfolgte danach die Codierung der als schwierig bzw. nicht codierbar eingestufteten Angaben. Insgesamt konnten 8.782 Angaben zum Wunschberuf codiert werden. Neben der bereits vorhandenen Missingkategorie -5 „wusste nicht, was ich werden wollte“ wurde die Kategorie -4 „Berufsangabe zu ungenau“ zusätzlich eingeführt.

Die Codierer, die die Codierung der Wunschberufe durchgeführt haben, waren bereits im Vorfeld als Codierer der Angaben zu beruflichen Tätigkeit eingesetzt, so dass sich die vor der Codierung des Wunschberufs durchgeführte Schulung primär

auf die Vereinfachungen der Codierregeln bezogen. Bei dieser Angabe haben wir auf die Ermittlung der Inter-Coder-Reliabilität verzichtet, gehen aber davon aus, dass sie eher über dem Wert in den ausgeübten Tätigkeiten liegt.

3 Codierung der Angaben zu Ausbildungen

In ALWA wurden Ausbildungen in einem vom Erwerbsmodul separaten Ausbildungsmodul erfasst. Unter Ausbildungen werden dabei verschiedene Arten von beruflichen Ausbildungen verstanden. Dazu zählen auf der einen Seite Lehre, Fach- oder Berufsfachschulausbildung, Hochschul-, Universitäts- oder Fachhochschulstudium. Auf der anderen Seite zählen dazu aber auch Umschulungen, Meister- oder Technikerbildungen, Lehrgänge bei einem Verband oder einer Kammer, wie z.B. IHK-Lehrgänge, Ausbildungen zur Vorbereitung auf die Laufbahnprüfung im öffentlichen Dienst, aber auch Promotion, Habilitation und Facharztzubereitung. Nicht in diese Kategorie fallen aber schulische Ausbildungen wie der Besuch des Gymnasiums, der Realschule oder der Hauptschule, diese wurden separat in einem Schulmodul erfasst. Ebenfalls nicht dazu gehören Anlernzeiten, Referendariate, Volontariate, das Anerkennungsjahr, Traineestellen, Vikariate, Praktika, pharmazeutische Praktika und Zeiten als Arzt im Praktikum. Diese wurden als Erwerbstätigkeiten erfasst. Kurze berufsbezogene Fort- und Weiterbildungen gehören ebenfalls nicht zur Kategorie Ausbildung.

Im Folgenden wird jedoch nur auf die Besonderheiten des Vorgehens bei der Codierung dieser Angaben eingegangen, da sie lediglich eine vereinfachte Adaption der beschriebenen Vorgehensweise zur Codierung der offenen Angaben zur beruflichen Tätigkeit darstellt.

3.1 Codierung der Ausbildungen ohne Kurse

Im Gegensatz zur beruflichen Tätigkeit wurde die Angaben zur Ausbildung mit lediglich einer offenen Frage erhoben, die je nach Typ der zugehörigen Ausbildung variiert wurde („In welchem Beruf genau haben Sie diese Ausbildung gemacht?“, „Welches erste Hauptfach haben Sie studiert?“, „In welchem Fach haben Sie promoviert / habilitiert?“, „Welchen Facharzt haben Sie gemacht?“ bzw. „Was war der Inhalt des Kurses/Lehrgangs?“). Für die Codierung der Ausbildungsberufe wurde ein Verfahren analog dem Verfahren zur Codierung der Angaben zur beruflichen Tätigkeit verwendet. Da nur eine Angabe codiert werden musste, vereinfachte sich das Verfahren erheblich.

In einem ersten Schritt erfolgte eine automatische Codierung mit Hilfe der DKZ, analog zu dem in 1.2 beschriebenen Verfahren. Die Vereinfachung bestand darin, dass keine Entscheidung zwischen zwei Treffern notwendig war, da pro Angabe nur ein exakter Treffer möglich war. Im nächsten Schritt erfolgte eine Codierung

der automatisch nicht codierbaren Treffer durch Codierer, erneut mit der Vereinfachung, dass keine Betrachtung von zwei Angaben notwendig war. In einem dritten Schritt erfolgte analog eine Codierung der in Schritt zwei als schwierig bzw. nicht codierbar eingestuften Angaben. Tabelle 4 zeigt die Ergebnisse der Codierung der Ausbildungsberufe in den jeweiligen Schritten.

Tabelle 4: Ergebnis der Codierung der Angaben zur Ausbildung

	N	Anteil
Angaben zur Ausbildung	15.091	100 %
Automatische Codierung	8.078	54 %
Codierung der Angaben ohne Treffer	7.013	46 %
Eindeutig zuordenbar	6.043	40 %
Problemfall	970	6 %

Auch hier kamen nur Codierer zum Einsatz, die bereits die Codierung der beruflichen Tätigkeiten durchgeführt haben. Die vor Beginn der Codierarbeiten durchgeführte Schulung bezog sich deshalb auch vordergründig auf die Vereinfachungen der Codierregeln.

Der Test der Inter-Coder-Reliabilität zeigt, dass die Ausbildungen zu einem höheren Grad gleichartig codiert wurden: Auf der 7-Steller-Ebene stimmten 62 % der Codes für die Ausbildungen zwischen den Codierern überein (vgl. Tabelle 5).

Tabelle 5: Ergebnisse der Inter-Coder-Reliabilitäten

Reliabilität	Codierung der offenen Angaben zu Ausbildungen
KldB 2010 - 7-Steller Ebene	62 %
KldB 2010 - 4-Steller Ebene	75 %
KldB 2010 - 2-Steller Ebene	84 %
ISCO 1988 – 4-Steller-Ebene	78 %

3.2 Codierung der Kurse

Kurse für den Erwerb von anerkannten Lizenzen, wie z.B. Schweißer-, Taxifahrer-, Gabelstaplerschein oder Netzwerkadministrator wurden ebenfalls im Ausbildungsmodul erfasst. Für diese Kurse existiert keine einheitliche nationale oder internationale Klassifikation. Allerdings wird im Rahmen des Vermittlungsprozesses in der Bundesagentur für Arbeit (BA) ein sogenannter Kompetenzkatalog eingesetzt, der auch für die Öffentlichkeit zugänglich ist.⁴ Dieser umfasst die Inhalte aller Berufsbeschreibungen im BERUFENET (<http://berufenet.arbeitsagentur.de>), die am Sei-

⁴ Der Kompetenzkatalog ist nach einer Registrierung unter folgendem Link zu finden: <http://download-portal.arbeitsagentur.de/files/hilfeDetail.do?ai=2516>.

tenende unter den Berufsinformationen „Kompetenzen“ beschriebenen Fertigkeiten verlinkt sind. Mit Hilfe dieses Kataloges wurden die Angaben in eine vorhandene Systematik eingeordnet. Allerdings hat sich im Verlauf dieses Prozesses gezeigt, dass für eine empirische Verwertung der codierten Angaben eine deutlich weniger differenzierte Kategorisierung ausreicht.

4 Codierung der Branchen

Die Angaben zu den Branchen bzw. Wirtschaftszweigen wurden in ALWA an zwei Stellen mit Hilfe einer offenen Frage („Zu welcher Branche bzw. zu welchem Wirtschaftszweig gehört Ihr Unternehmen?“) erfasst. Zum einen wurde die Branchenangabe für alle Erwerbsepisoden erhoben und zum anderen für Ausbildungen, die im Rahmen einer Lehre, d.h. Facharbeiterausbildung oder dualen Berufsausbildung gemacht wurden. Sowohl die Angaben der Beschäftigungs- als auch die Ausbildungsbranche wurden schematisch einheitlich nach der offiziell in Deutschland verwendeten Brancheneinordnung nach der Klassifikation der Wirtschaftszweige (WZ08) codiert.

Im Folgenden wird der Gesamtprozess der Codierung der Branchenangaben anhand der einzelnen Schritte dargestellt. Dieser Prozess unterteilt sich in drei Schritte. Der erste Schritt umfasste die Codierung der offen erhobenen Branchenangaben anhand einer Stichwortsuche in der WZ08 Klassifikation. Im zweiten Schritt wurden die nicht codierbaren Fälle mit Hilfe eines Abgleichs mit Sekundärinformationen wie z.B. Informationen aus Erwerbs- oder Ausbildungsverläufen der Person betrachtet und dann klassifiziert. Im dritten und letzten Schritt wurden für die weiterhin nicht codierbaren Fälle neue Konventionen aufgestellt. Zur Qualitätssicherung wurden ein Teil der Branchenangaben doppelt codiert und die Übereinstimmung der Codierungen überprüft.

4.1 Die Klassifikation der Wirtschaftszweige nach der WZ08

Die Klassifikation der Wirtschaftszweige, Ausgabe 2008 (WZ08) ist das offizielle Schema der Einteilung der Branchen und Wirtschaftszweige, das auch von der amtlichen Statistik in Deutschland verwendet wird.⁵ Diese Klassifikation erfasst alle gängigen Branchen bzw. Wirtschaftszweige, die in Deutschland wirtschaftliche Relevanz haben. Die Klassifikation der Branchen orientiert sich an der International

⁵ Vgl. Informationen auf den Seiten des Statistischen Bundesamtes:
<http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Content/Klassifikationen/GueterWirtschaftsklassifikationen/Content75/KlassifikationWZ08,template=renderPrint.psm>

Standard Industrial Classification (ISIC 4.1) bzw. der Adaption dieser Klassifikation auf europäischer Ebene, der Statistischen Systematik der Wirtschaftszweige in der Europäischen Gemeinschaft, NACE Rev.2 (Nomenclature des statistiques des activités économiques de la Communauté Européenne). Dadurch wird die nationale und internationale Vergleichbarkeit der Daten ermöglicht, da alle drei Systematiken auf der Ebene der Abteilungen, also auf Zweistellerebene, identisch sind.

Insgesamt besteht die WZ08 aus fünf Gliederungsebenen, für die maximal fünf Ziffern als Code vergeben werden.

Tabelle 6: Schema der WZ08 Einteilung:

Gliederungsebene	Anzahl	Code
Abschnitte (1. Stelle)	21	A-U
Abteilungen (2. Stelle)	88	01-99
Gruppen (3. Stelle)	272	01.1-99.0
Klassen (4. Stelle)	615	01.11-99.00
Unterklassen (5. Stelle)	839	01.11.0-99.00.0

Quelle: <http://www.destatis.de/>

Die ersten beiden Stellen der WZ08 lassen bereits grob die sektorale Zugehörigkeit erkennbar werden. Weitere Stellen ermöglichen eine Differenzierung nach Gruppen, Klassen und Unterklassen. Beispielsweise enthält der Zweisteller Angaben zur Zugehörigkeit eines Betriebes zur Landwirtschaft (01), der Herstellung von chemischen Erzeugnissen (20) oder Werbung und Marktforschung (73). Auf der Ebene der Abteilungen bis zur Ebene der Unterklassen wären genauere Zuordnungen möglich, wie z.B. Herstellung von Büromöbeln (31.01.1) oder Errichtung von Fertigteilbauten (41.20.2). Aufgrund der unzureichenden Angaben der Befragten, die sich meist auf der Zweistellerebene bewegen (z.B. Angaben wie „Landwirtschaft“, „Einzelhandel“, „Autohandel und Werkstatt“ etc.), war eine genauere Codierung oft nicht möglich, so dass durchgehend auf der Ebene der Abteilungen codiert wurde.

4.2 Die Codierung der Branchenangaben

4.2.1 Manuelle Codierung ohne Hilfsangaben

Die offenen Angaben zur Branche wurden ohne weitere Informationen (z.B. Angabe zu Erwerbs- und Bildungsverläufen) unter Berücksichtigung der WZ08 auf Zweistellerebene codiert. In einem ersten Schritt erfolgte eine manuelle Codierung,

d.h. ohne Zuhilfenahme eines automatischen Codierungsprozesses.⁶ Als Datengrundlage wurden die Suchmaschine des Bayerischen Landesamtes für Statistik und Datenverarbeitung in München⁷, die sich an der Struktur der WZ08 orientiert und die Gliederung der Klassifikation der Wirtschaftszweige, Ausgabe 2008 (WZ 2008)⁸ des Statistischen Bundesamtes Deutschland verwendet.

Folgende allgemeine Regeln wurden für die manuelle Codierung festgelegt:

1. Codiert wird grundsätzlich der in der offenen Angabe zur Branche enthaltene Begriff bzw. Teile des Begriffes, wenn er in der Suchmaschine bzw. in der Gliederung der Klassifikation der Wirtschaftszweige aufgeführt ist.
2. Bei mehreren Angaben, bei denen eine die andere näher spezifiziert, wird die jeweils spezifischere Angabe codiert.
3. Bei mehreren Angaben, bei denen die eine die andere nicht näher spezifiziert und es sich um unterschiedliche Angaben handelt, wird die zuerst genannte Angabe codiert (Geis, 2009, S.7).
4. Wird nur ein Treffer gefunden, wird dieser als Zweisteller übernommen, wenn nicht wurden Alternativen, Teile des Begriffes o.ä. als Suchbegriff eingegeben. Wenn hierbei nur ein Treffer gefunden wurde, wurde dieser übernommen.
5. Werden mehrere Treffer gefunden, wird der Fall an den zweiten Schritt, der Codierung unter Zuhilfenahme der Bildungs- und Erwerbsverläufe, übergeben. Ebenso wird mit Fällen ohne Treffer verfahren.

Für diesen wie für die folgenden Schritte der Codierung sind keine Angaben über die Anzahl bzw. Prozentanteile der codierten bzw. nicht codierbaren Fälle verfügbar, da sich die einzelnen Schritte teilweise überlappen und somit eine eindeutige Zuordnung einzelner Fälle nicht möglich ist. Im ersten Schritt konnte aber bereits

⁶ Eine automatische Vercodung wurde im Vorfeld versucht, führte jedoch zu keinem zufriedenstellenden Ergebnis. Nur ca. 5 % der Angaben hätten durch diesen Schritt codiert werden können. Bei einer visuellen Inspektion stellte sich heraus, dass die Angaben, die Personen über ihre Beschäftigungsbranche machen, sich deutlich von der Art und Weise unterscheiden, wie die Branchenangaben in der WZ 2008 klassifiziert werden.

⁷ Diese Suchmaschine ist unter:

<http://w3gewan.bayern.de/klassifikationen/klassw08/wz0801.htm> einzusehen.

⁸ Weitere Informationen unter:

<http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Content/Klassifikationen/GueterWirtschaftsklassifikationen/Content75/KlassifikationWZ08,templated=renderPrint.psm1>

ein Großteil der Branchenangaben codiert werden. Dennoch ergaben sich Nennungen der Branchenangaben der Befragten, die auch in dieser Phase nicht codiert werden konnten. Insbesondere waren Mehrfachtreffer zu behandeln (siehe Tabelle 7).

Tabelle 7: Mehrfachtreffer

	mögliche Zweisteller
Holzbranche, Holz u.ä.	02, 16, 31
Metall	24, 25, 28
Auto	29, 45, 77
Bergbau	05, 06, 07, 08, 09
Nahrungsmittel	10, 46, 47

Was sollte man also tun, wenn den offenen Branchenangaben verschiedene Codes zugeordnet werden können. Darüber hinaus waren Lösungen für jene Fällen zu definieren, für die die Branchenangaben so unspezifisch waren, dass sie im ersten Schritt nicht codiert werden konnten.

4.2.2 Manuelle Codierung mit Hilfsangaben

In einem zweiten Schritt wurden die noch nicht codierten Nennungen der Branchen mit den Bildungs- und Erwerbsverläufen der Befragten abgeglichen. Damit konnten beispielsweise Angaben wie „Dienstleistung“ oder „Bergbau“ eingeordnet werden oder zwischen einer Tätigkeit im Handel oder in der Produktion unterschieden werden. Ein solches Vorgehen ist ein Novum im Bereich der Codierung offener Angaben, da die Berücksichtigung von Sekundärinformationen bisher nicht als Methode implementiert wurde. Beispiele für den sinnvollen Einsatz der Codierung unter Zuhilfenahme der Bildungs- und Erwerbsverläufe sind in Tabelle 8 genannt.

Tabelle 8: Codierung mit Hilfsangaben (Einzelfälle), Auswahl

offene Angabe	Erster Schritt ohne Bildungs- und Erwerbsdaten	Zweiter Schritt mit Bildungs- und Erwerbsdaten
<i>Bergbau</i>	05 Kohlebergbau 06 Gewinnung von Erdöl und Erdgas 07 Erzbergbau 08 sonstiger Bergbau, etc.	z.B. aus den Erwerbsdaten ersichtlich: 05 Braunkohlebergbau
<i>Dienstleistung</i>	nicht codierbar, da zu viele Angaben in Frage kommen	z.B. aus den Erwerbsdaten ersichtlich: 63 Informationsdienstleistungen

In diesem Schritt wurde zusätzlich folgende allgemeine Regel formuliert:

6. Wenn eine Angabe mehreren zweistelligen Kennziffern zugeordnet werden kann, wird diejenige Angabe codiert, die empirisch am häufigsten vertreten ist.

4.2.3 Zuordnungsentscheidungen

Auch unter Zuhilfenahme von zusätzlichen Hilfsangaben konnten einige Angaben nicht codiert werden. Deshalb wurden in einem dritten Schritt generelle Zuordnungsentscheidungen getroffen:

1. Für die Angabe „Bergbau“ wird der Ort an dem die Erwerbstätigkeit ausgeübt wurde betrachtet, um Rückschlüsse auf die Art des Bergbaus zu ziehen (etort, Ort der Erwerbstätigkeit = Ruhrgebiet o.ä. typische Bergbaugebiete, die unter den Zweisteller 05 fallen -> Bergbau = 05; bei sonstigen Orten der Erwerbstätigkeit im Bergbau = 08).
2. Für die Angaben „Industrie“ und „Produktion“ in Verbindung mit unzureichenden Informationen aus den Bildungs- und Erwerbsverläufen wurde die neue Klasse 34 (Industrie, Produktion und Herstellung n.z.o.) gebildet.
3. Für die Angabe „Bau, Baugewerbe, Hoch- Tiefbau“, etc. ohne weitere verfügbare Informationen aus den Erwerbs- und Bildungsverläufen wird eine neue Abteilung 40 gebildet. Die darauffolgenden Klassen 41 - 43 beinhalten Branchenangaben des Baugewerbes.
4. Für die Angabe „Dienstleistung“ ohne genauere Informationen aus den Erwerbs- und Bildungsdaten wird die neue Abteilung 89 (Sonstige Dienstleistungen n.z.o.) geschaffen. Diese gehört zum Abschnitt der Dienstleistungen.
5. Die Angabe „Maurer“ wird als 41 (dem Hochbau zugehörig codiert) da anzunehmen ist, dass die Tätigkeit in dieser Branche empirisch am häufigsten vorkommt (Regel 6).
6. Die Angabe „Metall“ wird dem Zweisteller 25 „Herstellung von Metallerzeugnissen“ zugeordnet, obwohl grundsätzlich auch 24 „Metallerzeugung- und -bearbeitung“ bzw. 28 „Maschinenbau“ in Frage kommen. Hier wird ebenfalls Regel 6 verwendet.
7. Die Angabe „Handwerk“ ohne weitere Informationen aus den Erwerbs- und Bildungsverläufen wird mit 43 (sonstiges Ausbaugewerbe) codiert.
8. Die Angabe „Handel“ wird mit 47 (Einzelhandel) codiert, da empirisch häufiger der Einzelhandel in Verbindung mit Handel vorhanden ist (Regel 6).
9. Die Angabe „Groß- und Einzelhandel“ wird mit 46 (Großhandel) codiert, da die Wertigkeit des Großhandels nicht verloren gehen soll, zumal auch ein Großhandelsbetrieb einen angeschlossenen Einzelhandel haben kann.

Darüber hinaus wurden anhand der offenen Branchenangaben Zuordnungen vorgenommen. Die folgende Tabelle enthält solche Beispiele für Branchenangaben, die bei Betrachtung der offenen Angaben dennoch zugeordnet werden konnten.

Tabelle 9: Problembehaftete Fallbeispiele des dritten Schritts der Codierung

Angabe		Entscheidung	Begründung
a) Nennung einer Erwerbstätigkeit anstelle der Branche	Zeitungen austragen	53 Post- Kurier- und Expressdienste	Der Charakter des Austragens von Zeitungen ist entscheidend für die Einstufung der Branche in Post- Kurier- und Expressdienste.
	Hosen verkaufen	47 Einzelhandel	Die angegebene Tätigkeit findet definitiv im Einzelhandel statt, so dass hier mit 47 codiert wird.
b) ungenaue Angaben bei der Nennung der Branche	Hotel und Gaststätten	55 (Beherbergung)	Hier ist der Begriff Hotel ausschlaggebend, es wird nicht 56 (Gastronomie) codiert, da zu einem Hotelbetrieb gewöhnlich auch die Verpflegung der Gäste gehört. Das schließt unter Beherbergung auch die Gaststätten ein.
c) weitere Sonderfälle	Nennung des Unternehmensnamens: MC Donalds	56 (Gastronomie)	Es gibt Fälle, bei denen der Name des Unternehmens anstelle der Branche angegeben wird. Hier wird die Branche codiert, die hauptsächlich in Verbindung mit dem Unternehmen steht. Im konkreten Fall handelt es sich um ein Unternehmen im Bereich Gastronomie

In Tabelle A im Anhang finden sich weitere Codierungsbeispiele für den zweiten und dritten Schritt der Branchencodierung.

Insgesamt konnten von den 7.182 offenen Branchenangaben zum Ausbildungsbetrieb nur 4 nicht codiert werden (48 waren Missingangaben). Von den 32.956 Branchenangaben in den Erwerbsepisoden konnten (bei 32 Missings) nur 141 nicht codiert werden. Die nicht codierbaren Angaben umfassten hauptsächlich Mehrfachangaben und/oder ungenaue Angaben in der Nennung der Branche (z.B. „Privatwirtschaft“, „Büro“, „Radiosender“, „Pharmaunternehmen“) in Verbindung mit unzureichenden Angaben in den Erwerbs- und Ausbildungsverläufe.

Für die Codierung der Branchenangaben wurden Codierer eingesetzt, die bereits bei der Codierung der Angaben zur beruflichen Tätigkeit, zum Wunschberuf oder zum Ausbildungsberuf eingesetzt waren. Die Codierer erhielten eine Schulung zur Klassifikation der Wirtschaftszweige, zum Codierprozess der Branchenangaben und waren an der Entwicklung des dritten Codierschrittes beteiligt.

Zur Qualitätssicherung wurden alle Branchenangaben zur Erwerbstätigkeit mit einer Zweitcodierung abgeglichen. Hier beträgt diese Inter-Coder Reliabilität 73 %, d.h. in fast drei Viertel aller Fälle führte die Erst- und Zweitcodierung zu einem identischen Ergebnis auf Zweistellerebene. Bei der Zweitcodierung wurde allerdings nur der erste Schritt der Codierung reproduziert und nicht auf Angaben aus

der Erwerbs- und Ausbildungsgeschichte zurückgegriffen, so dass wir aus diesem Grund die Übereinstimmung für ebenfalls zufriedenstellend erachten.

5 Fazit

Ziel der Codierung war es, die offen erhobenen Angaben zur beruflichen Tätigkeit, Wunschberuf, Ausbildung und Branche in gängige Klassifikationen einzuordnen. Für die Codierung der offen erhobenen Angaben wurden jeweils eigene Codierverfahren für die Berufangaben und für die Branchenangaben neu entwickelt und erprobt. Insgesamt haben sich beide Verfahren als grundsätzlich geeignet erwiesen, um offen erhobene Angaben aus Befragen zu codieren. Der besondere Vorteil einer Transferierung der offenen Berufsangaben in DKZ-Siebensteller liegt darin, dass diese Codes ohne größeren Aufwand mit Hilfe eines Umsteigeschlüssels in verschiedene Berufsklassifikationen (z.B. ISCO) umgewandelt werden können.

Die niedrige Inter-Coder-Reliabilität ist jedoch nicht als Zeichen für „schlechte Codiererleistung“ zu werten, sondern zeigt eher auf, dass die bisherigen Berufsklassifikationen (KldB 1988, 1992) nicht geeignet sind, analytisch vergleichbare berufliche Einheiten zu erzeugen. Andererseits belegt sie auch, dass die DKZ teilweise zu differenziert ist, um als geeignete Basis für eine Berufscodierung benutzt werden zu können. Auch in der DKZ gibt es sehr viele Zuordnungsprobleme, die durch die neu erarbeitete Klassifikation der Berufe (vgl. Bundesagentur für Arbeit 2011a, 2011b) allerdings behoben sein dürften. Da bei der Entwicklung der KldB 2010 besonderer Wert auf die Schaffung analytisch vergleichbarer beruflicher Einheiten gelegt wurde und klare Zuordnungsregeln formuliert wurden, gehen wir davon aus, dass eine zukünftige Verwendung dieser Berufsklassifikation als Grundlage für eine Berufscodierung unter sonst gleicher Verfahrensweise, inklusive Codiererschulung, zu einer deutlich besseren Inter-Coder-Reliabilität führt.

Die Ergebnisse weisen zudem darauf hin, dass man darüber nachdenken sollte, in Zukunft von einer offenen Erhebung von Berufs-, Ausbildungs- oder Branchenangaben zugunsten der Verwendung dynamischer Datenbanken zur direkten Codierung von Berufs- oder Branchenangaben abzusehen. Durch unmittelbaren Einsatz solcher Datenbanken während des Interviews würden sie es dem Interviewer ermöglichen, die „richtigen“ Nachfragen zu stellen, um einerseits reliablere Angaben zu erzeugen und andererseits auch den Codieraufwand deutlich zu reduzieren.

Literaturverzeichnis

Antoni, M.; Drasch, K.; Kleinert, C.; Matthes, B.; Ruland, M.; Trahms, A.; (2010): Arbeiten und Lernen im Wandel. Teil I: Überblick über die Studie. (FDZ Methodenreport, 05/2010). Nürnberg.

Bundesagentur für Arbeit (Hrsg.) (2011a): Klassifikation der Berufe 2010. Bd. 1: Systematischer und alphabetischer Teil mit Erläuterungen. Nürnberg.

Bundesagentur für Arbeit (Hrsg.) (2011b): Klassifikation der Berufe 2010. Bd. 2: Definitorischer und beschreibender Teil. Nürnberg.

Ganzeboom, H. B. G.; de Graaf, P. M.; Treiman, D. J.; de Leeuw, J. (1992): A standard international socio-economic index of occupational status. *Social Science Research* 21, S. 1–56.

Geis, A. J. (2009): Handbuch für die Berufsvercodung. Mannheim: Gesis. Center for Survey Design and Methodology (Download 17.04.2012: http://www.gesis.org/fileadmin/upload/dienstleistung/tools_standards/hdb_0903.pdf)

Hoffmeyer-Zlotnik, J.; Geis, A. J. (2003): Berufsklassifikationen und Messung des beruflichen Status/Prestige. *ZUMA-Nachrichten* 27, Nr. 52, S. 125–138.

Hartmann, J.; Schütz, G. (2002): Die Klassifizierung der Berufe und der Wirtschaftszweige im Sozio-oekonomischen Panel. Neuvercodung der Daten 1984-2001. München: Infratest Sozialforschung.

International Labour Organisation (1990): International Standard Classification of Occupations: ISCO-88. Geneva: ILO.

International Labour Organisation (2008): Updating ISCO-88. (Download 28.01.2008 <http://www.ilo.org/public/english/bureau/stat/isco/isco88/intro.htm>)

Kleinert, C.; Matthes, B.; Antoni, M.; Drasch, K.; Ruland, M.; Trahms, A.; (2011): ALWA - New life course data for Germany. In: *Schmollers Jahrbuch. Zeitschrift für Wirtschafts- und Sozialwissenschaften*, Jg. 131, H. 4, S. 625-634.

Paulus, W.; Schweitzer, R.; Wiemer, S. (2010): Klassifikation der Berufe 2010. Entwicklung und Ergebnis. (Methodenbericht der Statistik der BA), Nürnberg.

Wegener, B. (1985): Gibt es Sozialprestige? *Zeitschrift für Soziologie* 14, Nr. 3, S. 209–235.

Anhang

Tabelle A: Codierungsbeispiele aus dem zweiten und dritten Schritt (Auswahl)

Nennung	Entscheidung	WZ 08
Internetunternehmen	63	Internetdienstleistungen
Dienstleistung	89	neue Klasse: Dienstleistung a.n.g.
Nonfood	46	Großhandel
Milchannahmestelle	1	Dienstleistungen in der Landwirtschaft
Medizin Technik	32	Herstellung von med. Apparaten
Kennzeichnungstechnik	28	Maschinenbau im Bereich Kennzeichnungstechnik
kaufmännisch	34	neue Klasse: Industrie / Produktion a.n.g.
Holdingsgesellschaft	89	neue Klasse: Dienstleistung a.n.g.
Hoch- u. Tiefbau	40	neue Klasse: Bau a.n.g.
Handwerksbetrieb	43	Handwerk empirisch am häufigsten Bauinstallation / sonst. Ausbaugewerbe
Handel	47	Einzelhandel empirisch am häufigsten bei Nennung "Handel"
Groß -und Einzelhandel	46	Großhandel empirisch am häufigsten bei Nennung "Groß- und Einzelhandel"
GGmbH	88	Gemeinnützigkeit = sonst. Sozialwesen
Getreideannahmestelle	1	Dienstleistungen in der Landwirtschaft
für einen selbständigen Unternehmer	82	Dienstleistung für ein Unternehmen / Firma
touristische einrichtung	55	hier: Hotel erkennbar
keinem	90	hier: Unterhaltungsdienstleistungen aus Sekundärdaten
Forst und landwirtschaft	2	hier: Forstwirtschaft aus Erwerbs- und Bildungsdaten erkennbar
Diensthandel	47	Handel = Einzelhandel
Volkseigener Betrieb	25	Betrieb im Bereich Metallverarbeitung
Überbetriebliche Ausbildungsstätte, die gefördert wird	85	Codiert als Unterricht, da Ausbildungsstätte
"kaufmännisch"	65	Bereich Versicherungen
Eloxall	20	Eloxieren ist chem. Verfahren, deswegen Chemie
Elektro u. Medien	47	Einzelhandel im Sinne eines Elektro-einzelhändlers

Impressum

FDZ-Methodenreport 04/2012

Herausgeber

Forschungsdatenzentrum (FDZ)
der Bundesagentur für Arbeit
im Institut für Arbeitsmarkt- und Berufsforschung
Regensburger Str. 104
90478 Nürnberg

Redaktion

Stefan Bender, Iris Dieterich

Technische Herstellung

Iris Dieterich

Rechte

Nachdruck - auch auszugsweise - nur mit
Genehmigung des FDZ gestattet

Bezugsmöglichkeit

http://doku.iab.de/fdz/reporte/2012/MR_04-12.pdf

Internet

<http://fdz.iab.de/>

Rückfragen zum Inhalt an:

Britta Matthes
Institut für Arbeitsmarkt und
Berufsforschung (IAB)
Regensburger Str. 104
90478 Nürnberg
Phone: 0911 / 179-3074
E-Mail: britta.matthes@iab.de