

Aufbau eines Metadaten- Informationsportals im Forschungsdatenzentrum des BIBB

- Auszüge des Antragstextes an das Bundesministerium für Bildung und Forschung -

Bundesinstitut für Berufsbildung
- Forschungsdatenzentrum -
Robert-Schuman-Platz 3
53175 Bonn

Telefon: 0228 / 107 - 20 41
Fax: 0228 / 107 - 20 20
E-Mail: fdz@bibb.de

www.bibb-fdz.de



Inhaltsverzeichnis

1	Ausgangslage	2
2	Ausbaubedarf der FDZ-Dokumentationen	3
	a. Erarbeitung und Etablierung eines inhaltlichen Metadatendokumentations-systems...	4
	b. Aufbau einer variablenorientierten Metadatendokumentation.....	4
	c. Stärkung der internationalen Bezüge	5
	d. Erschließung von Klassifikationssystemen auf der Datenebene.	5

1 Ausgangslage

Im Bundesinstitut für Berufsbildung wurde im Jahr 2008 ein Forschungsdatenzentrum (BIBB-FDZ) gegründet, dessen Aufgabe die Erschließung der Mikrodatenbestände des BIBB für die externe Berufsbildungsforschung ist. Alle Arten von einzelnen und Datensätze übergreifenden Informationen (so genannte Metadokumentationen) werden mit dem Regelbetrieb des BIBB-FDZ jedoch nicht abgedeckt. Dabei geht es nicht nur um inhaltliche und an einzelnen Variablen orientierte Dokumentationen, sondern auch um regionale Differenzierungen und international vergleichbare Klassifikationen des deutschen Berufsbildungssystems, für die eine Korrespondenz dieser Klassifikationen mit einzelnen oder kontextgebundenen Variablen in den BIBB-FDZ-Datensätzen zu erarbeiten ist.

Begriffe wie Metadaten oder Metadatenbanken (als Träger von Metainformationen) bezeichnen demnach kein eindeutiges Produkt oder eine entsprechende wissenschaftliche Methode/Vorgehensweise. In aller Regel sind Informationen über Mikrodaten gemeint, die auf verschiedenen Vertiefungsgraden (Ebenen) mit unterschiedlichen Produkten bereit gestellt werden. Im Wissenschaftsbereich ist es sinnvoll, mindestens drei Ebenen zu definieren:

- (i) an Datensätzen orientierte Metadatenbanken: Nach den Kriterien des Rats für Sozial- und Wirtschaftsdaten (RatSWD) erstellen FDZen Informationen über die ihnen vorliegenden Datenbestände. Insofern handelt es sich um eine FDZ-Regelaufgabe, die das BIBB-FDZ mit einem Online-Katalog umsetzt (vgl. <http://fdz.bibb.de/start.php>).
- (ii) an Themen orientierte Metadatenbanken und
- (iii) an Variablen orientierte Metadatenbanken.

Wie die drei Ebenen aufeinander aufbauen und dabei dennoch spezifische Produkte und Leistungen bereitstellen, lässt sich an einem Beispiel verdeutlichen: Metainformationen über Datensätze (i) müssen sich notwendigerweise auf einem allgemeinen Niveau bewegen und

Fallzahlen, Erhebungsmethode und vieles mehr darstellen. Damit lässt sich für wissenschaftliche Forschungsvorhaben abklären, ob es bereits prinzipiell geeignete Datenbestände gibt.

Als erste inhaltliche Annäherung geben thematische Stichworte, wie etwa „Ausbildung“, die Suchrichtung vor. Anschließend müssen Forschende klären, welche Aspekte in einem konkret vorliegenden Datenbestand behandelt werden (also etwa ob Ausbildungsbereitschaft lediglich über das Vorhandensein von Auszubildenden im Betrieb gemessen wird oder ob auch Verhaltens- oder Einstellungsfragen gestellt wurden und wenn ja, welche u.v.m.).

Bislang müssen sich Forschende diese Informationen in Papierform anhand der über FDZen bereitgestellten Fragebögen eigenständig erschließen, wobei die unterschiedlichen Herangehensweisen und Kompetenzniveaus der Forschenden zu individuell anderen Ergebnissen führen, die die Forschenden anschließend eigenverantwortlich und ohne Berufung auf einen Standard in der Fachöffentlichkeit vertreten müssen. Doch selbst wenn (i) bis (ii) zu einem positiven Ergebnis führten, ist die Sekundärverwertung der Daten noch nicht gesichert. Im nächsten Schritt sind Arbeiten an einzelnen Variablen oder auch Variablenblöcken notwendig. Wenn beispielsweise der Forschungsauftrag darin besteht zu prüfen, ob Jugendliche mit Migrationshintergrund bei der Ausbildungsbereitschaft von Betrieben diskriminiert werden, müssen sich Forschende selbstständig erschließen, ob und wenn ja, welche Messungen für den Migrationshintergrund in einem Datensatz, der sich hauptsächlich mit der Ausbildungsbereitschaft von Betrieben beschäftigt, verfügbar sind. Die relevanten Konstrukte zum Migrationshintergrund im Bereich Berufsbildungsforschung können über Verweise, etwa auf die KIBB-Wissenslandkarte und verwandte Produkte, recherchiert werden. Dort fehlt momentan allerdings noch der Rückbezug zu verfügbaren Datensätzen/Variablen im BIBB-FDZ. Da es verschiedene Definitionen von Migrationshintergrund gibt, die sich in der Regel an der Verfügbarkeit von Einzelinformationen in Datensätzen orientiert, müssen sich Forschende derzeit erst in solche, für die eigentliche Forschungsfragestellung randständige, Bereiche einarbeiten. Fehlen dazu Zeit, Ressourcen und/ oder eine Einarbeitungsbereitschaft, so wird die Auswertung im Extremfall nicht oder mit unverhältnismäßigem Aufwand durchgeführt. Außerdem müssen sich Forschende immer wieder individuell in solche und ähnliche Themengebiete einarbeiten.

2 Ausbaubedarf der FDZ-Dokumentationen

Im Zuge der Professionalisierung der Berufsbildungsforschung in Deutschland insgesamt und aufgrund technischer und organisatorisch-struktureller Verbesserungen der informationellen Infrastruktur ist es mittlerweile möglich, Informationen zu einzelnen Datensätzen auf einer übergeordneten Ebene zu strukturieren. Diese so genannten Metadaten-

Dokumentationen stehender wissenschaftlichen Forschung zu Recherchezwecken zur Verfügung. Derartige Informationen sind für die Berufsbildungsforschung wichtig, weil Kenntnisse über die jeweilige inhaltlich-methodische Leistungsfähigkeit von Sekundärdaten ausschlaggebend dafür sind, ob für Forschungsarbeiten Datenerhebungen erfolgen müssen oder nicht. Die bisherigen Erfahrungen des BIBB-FDZ belegen einen hohen externen Informations- und Beratungsbedarf über die Leistungsfähigkeit der BIBB-Datensätze für individuelle Forschungszwecke.

Für den Ausbau der FDZ-Dokumentationen sind im Einzelnen folgende Arbeiten erforderlich:

a. Erarbeitung und Etablierung eines inhaltlichen Metadatendokumentationssystems

Um die Gesamtheit der BIBB-Datenbestände in ihrem Auswertungspotenzial miteinander vergleichen zu können, ist die Erschließung von datensatzübergreifenden Informationen erforderlich. Hier geht es insbesondere darum, thematische Anker zu setzen und die BIBB-Datenbestände darin zu verorten. Um nicht in Konkurrenz zu bestehenden Berichtssystemen des BIBB zu treten und um Doppelarbeiten zu vermeiden, ist eine enge hausinterne Kooperation mit KIBB, respektive der Wissenslandkarte des BIBB, notwendig. Die ersten Sondierungsgespräche identifizierten insbesondere zwei Arbeitsstränge: Die einzelnen Komponenten der Wissenslandkarte für die FDZ-Zwecke sind neu zu ordnen, so dass eine Korrespondenz mit den BIBB-Mikrodatenbeständen (und entsprechenden Forschungsprojekten) erreicht wird. Diese konzeptionelle Aufgabe ist zu Beginn des Projekts zu leisten. Anschließend ist die Verschlagwortung bei der Wissenslandkarte und beim FDZ-Online-Katalog zu erweitern und so umzusetzen, dass Inhalte der Wissenslandkarte mit Informationen über einzelne Datensätze miteinander verknüpft werden. Wünschenswert ist, die Forschungsliteratur (insbesondere deren Katalogisierung in der Literaturlandkarte Berufliche Bildung, LdBB) mit in das inhaltliche Metadatendokumentationssystem einzubinden. Darüber hinaus besteht ein gesondertes Interesse darin, die zu entwickelnden Dokumentationssysteme mit anderen FDZ/Datenservicezentren (DSZ) zu vernetzen. Respektive im (Berufs-)Bildungsbereich ist abzusehen, dass eine gewisse Kooperation mit anderen FDZ/DSZ notwendig sein wird. Dies betrifft die FDZ des Statistischen Bundesamtes (Bildungsberichterstattung), das IQB (Schuldaten), das HIS (Hochschulbildung) sowie das DJI (Studien zu und über jungen Menschen), wobei das HIS und das DJI gerade beginnen ihre jeweiligen FDZ zu etablieren.

b. Aufbau einer variablenorientierten Metadatendokumentation

Derzeit steht die FDZ-Datenbank nur als Grundmodul zur Verfügung, in dem nicht zu detailreiche Informationen zusammengestellt werden. Diese sind mit konkreten Variablen und ihren Ausprägungen (Wertebereichen) zu verknüpfen, damit sich die Eignung der Daten für ein wissenschaftliches Forschungsvorhaben beurteilen lässt. Die Beurteilung kann für externe

Auswertungswünsche nur vom FDZ-Personal geleistet werden, da sie einen arbeitsintensiven Abgleich mit den Originaldaten erfordert. Bei einer (weiteren) Erhöhung der Anfragefrequenz werden dadurch so viele FDZ-Ressourcen gebunden, dass die Durchführung wichtiger Kernarbeiten darunter leidet.

Der Unterschied zu den Arbeiten für a) besteht darin, dass sich das beabsichtigte Dokumentationsmodul ausschließlich auf der Ebene der BIBB-Mikrodatensätze bewegt und Gemeinsamkeiten und Unterschiede zwischen inhaltlich zumindest ähnlichen Variablen in den einzelnen Datensätzen identifiziert und gemeinsam dokumentiert werden. Die Dokumentation soll nach Möglichkeit als Erweiterungsmodul der FDZ-Datenbank umgesetzt werden. Dabei ist es nicht ausreichend, lediglich die Inhalte der entsprechenden Variablen miteinander zu vergleichen und zu systematisieren, vielmehr sind auch deren jeweilige Wertebereiche vor dem Hintergrund der in den einzelnen Erhebungen erzielten Fallzahlen und methodischen Abdeckungsgrade zu berücksichtigen.

c. *Stärkung der internationalen Bezüge*

Neben einer englischsprachigen Datendokumentationen samt Variablenlabels, ist zu prüfen, inwiefern sich Variablen in einzelnen Datensätzen in international vergleichbare Klassifikationen überführen lassen. Eine solche Aktivität ist besonders arbeitsintensiv, weil es nicht nur teilweise konkurrierende internationale Klassifikationssysteme gibt (für Berufe etwa ISCO, ICSE, MPS, ESeC), sondern auch die konkreten Zuordnungen, aufgrund der Sonderstellung des deutschen Berufsbildungssystems im internationalen Kontext, bezüglich der „korrekten“ wissenschaftlichen Anwendbarkeit teilweise umstritten sind.

Für eine Stärkung der internationalen Bezüge müssen diese verschiedenen Perspektiven den BIBB-Mikrodaten gegenübergestellt und anschließend, möglichst in ihrer Gesamtheit, auf einzelne BIBB-Datensätze angewendet werden. Es ist zu berichten, wie sich in einzelnen Datensätzen – möglicherweise in Abhängigkeit von den Fallzahlen und zugrunde gelegten Variablen – empirische Ergebnisse bei der Anwendung der einen oder anderen internationalen (Standard-)Klassifikation ändern, worauf diese Veränderungen zurückzuführen und wie sie inhaltlich zu interpretieren sind. Zur Ergebnissicherung und als Dienstleistung für die Berufsbildungsforschung sind die entsprechenden Programmcodes frei verfügbar zu machen, damit sie von Anwendern selbstständig auf die BIBB-FDZ-Datensätze angewendet werden können.

d. *Erschließung von Klassifikationssystemen auf der Datenebene.*

Die im Bereich der Berufsbildungsforschung gängigen Klassifizierungssysteme sollen in Programmsyntax übersetzt und auf geeignete Datensätze angewendet werden. Die Bezüge zwischen einzelnen Klassifizierungssystemen sind zu erarbeiten. Im Unterschied zu c) geht es hierbei ausschließlich um nationale Klassifizierungssysteme, etwa im Bereich von Berufen

