

## **Workshop zur „Datengewinnungsstrategie für die Bildungsstatistik“ am 13.02.07 in Berlin**

**Rede von Staatssekretär Joachim Jacobi  
(Vorsitzender der Kommission für Statistik der Kultusministerkonferenz)**

### **Überblick über die Sachlage**

Herr Präsident, meine sehr verehrten Damen und Herren. Meine Aufgabe ist es zunächst darzulegen, wie der Stand der Überlegungen in den Gremien der KMK, insbesondere auch der Kommission für Statistik, in diesen Fragen ist.

Lassen Sie mich dazu mit der zunächst überraschenden Frage anfangen: Wozu braucht man überhaupt eine solche Schulstatistik?

Nun, die Antwort ist relativ einfach. Sie ist aber für das, was hier genau geplant wird, von entscheidender Bedeutung. Zunächst einmal brauchen diejenigen, die in den bildungspolitischen Fragen Konzepte entwickeln und Entscheidungen fällen, eine Basis für ihr administratives und politisches Handeln. Das, was man gelegentlich Steuerungswissen nennt, wird benötigt, um diese Entscheidungen sachgerecht fällen zu können. Ein solches Wissen kann man nur aus der Statistik gewinnen. Wenn man sich über den Zustand des Bildungswesens im Detail informieren will, braucht man dazu Daten, schlicht Statistik. Nicht zuletzt: Wer Schule und Unterricht wirksam in ihrer Qualität verbessern will, der muss Erkenntnisse darüber haben, wo Stärken und Schwächen liegen, wo genau mit Reformen anzusetzen ist – Bildungsstatistik ist hierzu unverzichtbar. Aber auch die Öffentlichkeit hat ein Recht darauf, zuverlässige Informationen über den Zustand des Schulwesens oder über den Zustand von Bildung allgemein in Deutschland zu bekommen. Das beweisen ja nicht zuletzt zahlreiche Parlamentsanfragen, denen sich alle Landesregierungen relativ regelmäßig ausgesetzt sehen, in denen häufig sehr detailliert und sehr speziell über einzelne Fragen Auskunft verlangt wird. Solche Parlamentsanfragen können nur auf der Basis verlässlicher Daten beantwortet werden. Auch dazu braucht man Statistik. Drittens, aber nicht zuletzt ist in Artikel 91b des Grundgesetzes seit Neuestem die Bildungsberichterstattung aufgenommen worden. Sie hat damit Verfassungsrang erlangt, auch das ist im Zusammenhang mit dem, was wir hier zu diskutieren haben, von Bedeutung.

Weil Bildung Ländersache ist, haben sich die Länder 1981 und erneut im Jahr 2000 zur Bereitstellung vergleichbarer Daten für die Bildungsstatistik verpflichtet. In den Ländern werden diese Daten erhoben, an das Statistische Bundesamt sowie zum Teil auch an das KMK-Sekretariat weitergeleitet, und von dort erfolgt dann die Weitergabe auch an internationale Institutionen, eine Aufgabe, die in den letzten Jahren immer wichtiger geworden ist.

Wie werden jetzt diese Daten gewonnen?

Lassen Sie mich einen Blick zurückwerfen in die Vergangenheit, die ja viele noch gut kennen. In der Vergangenheit war es üblich, dass die Schülerdaten auf Karteikarten in den Sekretariaten der Schulen notiert waren und die Fragen, die für die Statistik von Interesse waren, per Strichliste zusammengestellt wurden. Die einzige Möglichkeit, auf diese Weise zu Daten zu kommen, war, dass zu bestimmten Fragen die Antworten als Summen angegeben wurden. Die Frage „Wie viele Jungen oder wie viele Mädchen machen in einem Jahr in einem

Land Abitur?“ wurde dadurch beantwortet, dass die Schulen jeweils Zahlen zusammengetragen haben; diese wurden aufaddiert, und man konnte anschließend sagen: 52% der Abiturienten sind Mädchen und 48 % sind Jungen. Auf diese Weise konnte man, wenn konkret danach gefragt wurde, auch noch beantworten, wie viele Migranten in Deutschland das Abitur bestanden haben. Wenn man sich dann aber dafür interessiert hat, wie viele Jungen unter den Migranten das Abitur bestehen, dann war das aus diesen beiden Summenangaben nicht zu berechnen. Ich schildere dieses Beispiel, um deutlich zu machen, wie Statistik noch bis vor wenigen Jahren zu ihren Daten kam und welche Grenzen sich damit zwangsläufig aufboten.

Es gibt eine ganze Reihe von Fragen, die mit dieser Methode der Statistik offenkundig nicht beantwortbar sind. Die Frage „Wie wirksam ist das „Sitzenbleiben“?“, genauer: die Frage „Wie häufig bleiben Schüler, die einmal „sitzen geblieben“ sind, im Laufe ihrer Schulzeit ein zweites Mal sitzen und müssen dann die Schullaufbahn wechseln, wie viele machen dann welchen Abschluss?“ Diese Fragen sind auf der Basis solcher Zahlen offenkundig nicht zu beantworten. Wenn man das wissen wollte, müsste man eigene Fragen neu stellen und dann wieder per Strichliste auswerten. Natürlich ist das im Prinzip eine Möglichkeit. Aber dieses Verfahren ist teuer, es ist langsam, es ist unflexibel, es ist in hohem Maß bürokratisch und die Zahlen, die geliefert werden, sind häufig noch von schlechter Qualität. Denn auch das ist eine Erfahrung, auf die man in diesem Zusammenhang hinweisen muss: Je höher der Aufwand ist für die Stelle, die Auskunft geben muss, umso schlechter ist die Qualität der Daten, die gewonnen werden.

Diese Situation hat sich mit der Einführung von Schulverwaltungssoftware und der zum Teil schon erfolgten Umstellung der Schulstatistiken in den Ländern geändert. In den letzten Jahren ist in den meisten Ländern eine Schulverwaltungssoftware angeschafft worden, die zunächst den Sinn hat, die Verwaltungsabläufe in den Schulen zu vereinfachen. Wenn Sie so wollen: Festplatte statt Karteikarten. Das vereinfacht zunächst die Gewinnung der Daten, weil die Summen, die in der Vergangenheit per Strichliste gewonnen werden mussten, jetzt sehr viel einfacher aus dieser Verwaltungssoftware gewonnen werden können. Allerdings bietet die Nutzung dieser Software eine weitere Möglichkeit: Sie bietet die Möglichkeit, Individualdaten zu erheben, nicht mehr nur Summen zu erfassen. Selbstverständlich anonymisiert und selbstverständlich da, wo dieses Verfahren bereits eingeführt worden ist, immer in Absprache mit den jeweiligen Datenschutzbeauftragten. Alles, was ich Ihnen jetzt schildere, geschieht in den Ländern und wird natürlich in den Ländern mit den jeweils dafür zuständigen Datenschutzbeauftragten auch besprochen. Es ist offenkundig, dass eine Reihe von Fragen auf diese Weise besser und einfacher beantwortet werden kann.

Zunächst einmal ist es aber notwendig, dass sich die Länder darauf verständigen, welche Individualdaten erhoben werden sollen, um auf die von mir skizzierten Fragen Auskunft geben zu können. Die Antwort darauf ist der so häufig erwähnte Kerndatensatz. Der Kerndatensatz ist nichts anderes als eine Verständigung der Länder darüber, welche Individualdaten erhoben werden sollen. Eine solche Verständigung hat es in der Vergangenheit auch bei den per Kartei und Strichliste ermittelten Summendaten gegeben. Jetzt wird diese Verständigung auf der Basis von Individualdaten notwendig, damit die Statistiken, die in den jeweiligen Ländern gewonnen werden, auf Dauer vergleichbar sind. Das führte in der KMK im Jahr 2003 zu dem grundsätzlichen Beschluss, Länderstatistiken zukünftig auf anonymisierte Einzeldatensätze umzustellen und einen gemeinsamen Kerndatensatz zu finden. Beauftragt wurde damit der damalige Unterausschuss Schulstatistik, der jetzt in die Kommission für Statistik der KMK übergegangen ist. Es ist ein Gremium der

KMK, das sich mit Detailfragen dieser Aufgabe auseinanderzusetzen hat. Dieser Grundsatzbeschluss aus dem Jahr 2003 wurde dann 2005 und 2006 von der Amtschefkommission bekräftigt und konkretisiert. Die Länder verpflichten sich, den gemeinsamen Kerndatensatz einzuführen. Die öffentlichen Schulen sollen ab dem Schuljahr 2008/2009 auf der Basis dieses Kerndatensatzes die entsprechenden Daten liefern. Bis dahin wird jährlich in der KMK ein Bericht über den Stand der Einführung abgegeben. Und selbstverständlich ist klar, dass - wegen der verfassungsmäßig deutlich formulierten Zuständigkeit - die Länder dafür die rechtlichen und sächlichen Voraussetzungen schaffen müssen.

Der Kerndatensatz enthält anonymisierte Einzeldatensätze zu einer Reihe von Bereichen. Er erhält einen Datensatz zur Schule, einen Datensatz zur Klasse, er erhält – das steht häufig im Zentrum der Diskussion – einen Datensatz zu jedem Schüler, für die Unterrichtseinheiten, zu den Absolventen und Abgängern und auch zu den Lehrkräften und den Lehrerbewegungen. Lassen Sie mich darauf hinweisen, dass bei dem, was hier erhoben wird, nur sehr wenige Merkmale ermittelt werden, die nicht bereits schon in der Vergangenheit erhoben wurden. Es geht also nicht darum, dass eine Menge von neuen Daten gesammelt wird, sondern es geht um einen besseren Umgang mit diesen Daten. Viele Verwaltungsdaten, die zu Zwecken der Verwaltung in den Schulen vorhanden sind, werden für die Statistik nicht benötigt und deswegen auch nicht in statistische Daten überführt. Denn die Statistik interessiert sich nicht für den Einzelfall. Die Statistik interessiert sich für Strukturen und Prozesse und nicht dafür, wie ein einzelner individueller Schüler etwa seine Schullaufbahn durchläuft. Es werden keine Telefonnummern, keine Namen, keine Personalnummern – um Beispiele zu nennen - in die Statistik überführt und benutzt.

Lassen Sie mich in diesem Zusammenhang ein Wort zu der ID-Nummer sagen, die ja in der öffentlichen Diskussion vor einiger Zeit eine große Rolle gespielt hat. An dieser Stelle der Diskussion ist die ID-Nummer nichts anderes als ein Hilfsmerkmal, um die Datensätze auseinanderzuhalten. Technisch gesprochen ein Primärschlüssel für die Datenbank – und weniger technisch gesprochen: ein Hilfsmerkmal, um zwei Schüler, die in sehr vielem übereinstimmen, etwa Zwillinge, die die gleiche Schule besuchen und in die gleiche Klasse gehen, trotzdem als individuelle, unterschiedliche Datensätze identifizieren zu können. Sie können dafür die Schüler schlicht durchnummerieren oder auf eine sonst beliebige Art und Weise eine Unterscheidung herbeiführen. Das ist an dieser Stelle die ID-Nummer. Das ist die aktuelle Beschlusslage der KMK.

Der Kerndatensatz soll Informationen über Entwicklungen im Bildungswesen gewinnen helfen, vergleichbare Daten innerhalb Deutschlands, Grundlagen für Prognosen und Modelle, es soll die Überprüfung der Zielerreichung sicherstellen. Es soll auch in internationalen Statistiken sicherstellen, dass Deutschland korrekt dargestellt wird. Lassen Sie mich darauf hinweisen, dass die KMK damit auch keinen Sonderweg geht. Andere sind schon genauso weit oder weiter. Einzeldatensätze sind beschlossen in der Kinder- und Jugendhilfestatistik, in der Hochschulstatistik, in der Berufsbildungsstatistik und in der großen Mehrzahl der europäischen Nachbarländer oder sogar schon Praxis.

Neben dem Kerndatensatz gibt es aber eine weitere Überlegung. Parallel zu dieser Entwicklung ist der Bedarf an Informationen in den letzten Jahren offenkundig gestiegen. Die internationalen Vergleiche nehmen zu. Es gibt viele Anfragen, die in der Vergangenheit nicht gestellt worden sind. Deswegen hat sich das Konsortium für die nationale Bildungsberichterstattung dazu geäußert, wie man für eine verbesserte Berichterstattung eine

verbesserte Datenlage herbeiführen kann. Da gab es eine Reihe von Forderungen aus der empirischen Wissenschaft: Die Erhebung individueller Verlaufsdaten ist eine davon, die Darstellung erworbener Kompetenzen eine andere. Es ist auch die Frage diskutiert worden, ob die Darstellung des sozioökonomischen Hintergrunds der Schülerinnen und Schüler zu den Fragen gehört, die untersucht werden sollen. Das sind insoweit Forderungen aus der empirischen Wissenschaft. Die KMK hat daraufhin eine Arbeitsgruppe eingesetzt, die sich damit beschäftigen soll, diese Forderungen zu analysieren und zu bewerten. Diese Arbeitsgruppe hat sich für drei Maßnahmen mit höchster Priorität ausgesprochen.

Diese Maßnahmen möchte ich Ihnen gerne vorstellen und den Stand der Erörterungen in der KMK dazu mitteilen.

Punkt eins ist die Bekräftigung, Einzeldatensätze in der Statistik einzuführen. Das ist Beschlusslage der KMK und muss eigentlich nicht erneut thematisiert werden. Es ist nur noch einmal als ein besonders wichtiger Beitrag für eine verbesserte Bildungsberichterstattung erwähnt worden.

Der zweite Punkt ist die Möglichkeit, die vorhandenen Daten möglichst effektiv auswerten zu können. Die Länder sind die Herren über die Daten. Wir müssen aber viele Fragen bundesweit beantworten können. Wir brauchen effektive Auswertungsmöglichkeiten. Ein Weg, zu effektiver Auswertung zu kommen, ist die Einführung einer gemeinsamen Datenhaltung der Länder. Das kann zum Beispiel so funktionieren, dass für bestimmte Auswertungen – beileibe nicht für alle – ein Statistisches Landesamt im Auftrag der anderen - und in jedem Falle natürlich auch mit Zustimmung der anderen - die Daten auswertet. Möglicherweise kann man einem Statistischen Landesamt in einem definierten Umfang Zugriff auf solche Daten erlauben. Alles das ist im Moment noch in der Diskussion und wird unter bundesweiter Datenhaltung verstanden. Hier ist der Weg der technischen Realisierung noch nicht entschieden. Wichtig ist nur, dass wir zu einer Möglichkeit kommen, die Daten wirklich effektiv und dann auch bundesweit auswerten zu können.

Der dritte Punkt ist die Frage, wie kommen wir auf Dauer zu verlässlichen Möglichkeiten, Bildungsverläufe untersuchen zu können.

Lassen Sie mich das auch wieder an dem vorhin schon genannten Beispiel mit den Sitzenbleibern erläutern. Was man leicht beantworten kann, ist die Frage „Wie viele Schüler müssen im Durchschnitt eines Jahres eine Klassenstufe wiederholen?“. Dann kommen Sie von Land zu Land zu unterschiedlichen Ergebnissen. Ich nehme einmal an: 3 %. Wenn ich dann weiß, es waren in diesem Jahr 3 % und im letzten Jahr auch 3 %, kann ich aber die Frage „Wie oft geschieht das bei einem einzelnen Schüler hintereinander?“ schon nicht beantworten. Das ist mit Summendaten offensichtlich nicht möglich. Es ist aber auch mit Individualdaten nicht möglich, weil ich die Schüler, die „sitzen geblieben“ sind in diesem Schuljahr nicht in eine Beziehung setzen kann zu denjenigen Schülerinnen und Schülern, die im vergangenen Jahr „sitzen geblieben“ sind. Das geht nur, wenn es technisch möglich gemacht wird, dass die Sitzenbleiber dieses Jahres mit den Sitzenbleibern des vergangenen Jahres in Zusammenhang gebracht werden können und man die erstmaligen Wiederholer von denen unterscheiden kann, die bereits zum zweiten Mal „sitzen geblieben“ sind. Und das ist mit modernen Mitteln möglich, ohne dass die Namen dazu benötigt werden. Das ist auch möglich, ohne dass es eine eindeutige Zuordnung zwischen irgendeiner Nummer und irgendeinem Namen geben muss. Es ist machbar durch eine sogenannte Einwegverschlüsselung. Die Einwegverschlüsselung soll jedem Individuum, jedem Schüler, aus der Summe seiner Daten eine Kennzahl zuordnen, ohne dass es möglich ist, von der Kennung umgekehrt auf den Schüler zu schließen. Das „Zurückschließen“ auf den Schüler ist nicht nur deswegen unmöglich, weil man den Zugriff verhindert, sondern es gibt auch Wege,

dies technisch unmöglich zu machen, so dass der Datenschutz hierbei in vollem Umfang sichergestellt wird. Diese Einwegverschlüsselung wurde oft als Identifikationsnummer bezeichnet, und das hat zu vielen Missverständnissen geführt. Dabei geht es nur darum, in der Bildungsbiographie von Schülern Zeitreihen aufzustellen, um Fragen, wie ich sie genannt habe, zu beantworten. Der Rückschluss von solchen Zeitreihen auf einen bestimmten Schüler bleibt unmöglich.

Deutlich geworden ist damit auch, dass die KMK zu keinem Zeitpunkt ein Register geplant hat, das man sich ja so vorstellen könnte, dass etwa zehn Millionen Schüler alphabetisch aufgelistet werden und – ich darf es karikieren – mit Namen, Wohnort, Noten und mit allem Fehlverhalten aus diesem Register zu entnehmen sind. Nichts davon ist wahr und nichts davon war zu irgendeinem Zeitpunkt beabsichtigt. Der Begriff Register liegt meilenweit neben dem, was hier tatsächlich geplant worden ist. Durch die Einwegverschlüsselung ist die Rückführung der anonymisierten Daten auf den Einzelfall schlicht unmöglich.

Die Konzentration auf die genannten drei Punkte - Einführung des Kerndatensatzes, die Möglichkeit von Verlaufsaussagen und die Gewinnung effektiver Strategien zur Auswertung der Daten - erschien der Arbeitsgruppe, die sich mit der Datengewinnungsstrategie auseinandergesetzt hat, als ein brauchbarer Kompromiss, der einerseits nicht alle Anforderungen der Bildungsforschung beantwortet, aber andererseits nichts weniger als einen Quantensprung für die Statistik bedeutet, einen Quantensprung in der Frage, welche Anliegen der Bildungsberichterstattung zukünftig bedienbar sind. Dieser Vorschlag wird im Moment in den Gremien der KMK diskutiert. Hierzu gibt es noch keinen Beschluss. Zu dessen Vorbereitung dient dieser Workshop. Es ist wichtig, zu unterscheiden, dass wir grundsätzlich einig darüber sind, dass wir den Kerndatensatz einführen wollen. Aber die Frage, wie wir mit Verlaufsdaten und mit einer gemeinsamen Datenhaltung umgehen, ist in der KMK noch in der Diskussion.

Dabei wollen wir diese Entscheidungen in enger Absprache mit dem Datenschutz treffen. Deswegen hat es im Dezember 2006 ein erstes Gespräch mit Vertretern der Konferenz der Datenschutzbeauftragten gegeben, in dem wir den Arbeitsstand des Konzeptes diskutiert haben. Deshalb hat die KMK beschlossen, heute diesen Workshop hier stattfinden zu lassen, in dem ebenfalls mit Vertretern des Datenschutzes und mit Vertretern der empirischen Bildungsforschung darüber diskutiert wird, was notwendig, was möglich und was machbar ist.

Wichtig ist in diesem Zusammenhang, dass wir auf die strikte Trennung von Verwaltungsdaten der einzelnen Schulen und den von uns angestrebten Statistikdaten größten Wert gelegt haben. Die jetzt vorgesehenen Vorschläge sollen ermöglichen, dass Statistikdaten aus Verwaltungsdaten gewonnen werden, und zwar unter Wahrung der Anonymität der Schülerinnen und Schüler. Es ist gesichert, dass umgekehrt aus den Statistikdaten keine Verwaltungsdaten ermittelt werden können.

Es wird im Zusammenhang mit dem Datenschutz auf den Beschluss des Bundesverfassungsgerichts bei der Volkszählung zurückgegriffen. Lassen Sie mich auf einen wesentlichen Unterschied hinweisen. Bei der Volkszählung ging es darum, dass Daten zum Zwecke der Statistik erhoben wurden und dann sichergestellt werden musste, dass diese Daten nicht für andere Zwecke missbraucht werden. Davon kann hier keine Rede sein. Hier werden keine neuen Daten zum Zwecke der Statistik erhoben, sondern hier werden Daten erhoben zum Verwaltungshandeln der Schulen. Alle Datensätze, die in den statistischen Datenbanken anonym vorhanden sein werden, sind in den Verwaltungsdaten der Schule mit voller Kennung

des Namens bereits vorhanden. Mir hat auch bislang noch niemand plausibel machen können, warum durch den Vorgang der Anonymisierung und des Einstellens in eine anonyme Datenbank Datenschutzbelange gefährdet sein sollen, die nicht gefährdet waren, solange nur die Schule über diese Namen vollständig verfügte. Natürlich muss und wird in der Schule größten Wert auf den Datenschutz gelegt. Dafür muss in den Ländern gesorgt werden. Mir ist aber auch wichtig, dass dieser Zusammenhang deutlich wird: Die Daten über die Schüler sind in der Schule bereits vorhanden und werden durch die Anonymisierung und durch das Zuführen in die Statistik keineswegs unsicherer.

Lassen Sie mich abschließend nochmals betonen, dass es noch offene Fragen gibt: Offenkundig ist die Frage, wie man die gemeinsame Datenhaltung organisiert, noch nicht abschließend beantwortet. Deshalb ist es gut und richtig, in einem solchen Workshop diese Fragen zu diskutieren. Es ist richtig und wichtig, dass über das, was hier gemacht wird, auch kontrovers gestritten wird. Es ist uns, der KMK, wichtig: Es geht nicht um eine wilde Datensammelei, um ein ufer- und nutzloses Sammeln von Daten. Sondern es geht darum, die Fragen, die von der Öffentlichkeit gestellt werden, auf einer gesicherten Datenbasis beantworten zu können und der Bildungspolitik verlässliche Entscheidungsgrundlagen zu liefern. Das ist der Sinn des Workshops. Ich würde mich sehr freuen, wenn es gelänge, mit diesem Workshop die Diskussion auf sachliche Grundlagen zurückzuführen und nicht mit Unterstellungen und Vermutungen zu argumentieren, die in der KMK in dieser Weise nie vorgesehen waren. Ich bedanke mich für Ihre Aufmerksamkeit.