

Datenbanklösungen für die Sozialwissenschaften

Data Warehouse- und Business Intelligence- Ansätze im Nationalen Bildungspanel

5. KSWD, Wiesbaden

14.01.2011

David Schiller, M.A.

Gliederung

Einblicke in die Planungen des Nationalen Bildungspanels

1. Datenstruktur des Nationalen Bildungspanels (NEPS).
2. Grenzen normaler „Flatfiles“.
3. Möglichkeiten von Datenbanklösungen - Beispiele.
4. Ausblick.

Das Nationale Bildungspanel (NEPS)

- Längsschnittdaten für die empirische Bildungsforschung
- Fragebögen und Kompetenztests
- Informationen über
 - Zielpersonen (über 60.000)
 - Eltern
 - Lehrer/Erzieher
 - Leitung (Kindergarten/Schule)
 - Institutionen (Kindergarten/Schule/Hochschule)

LEITUNG UND KOORDINATION DES NEPS

IM INBIL BAMBERG

SÄULE 1

KOMPETENZ-
ENTWICKLUNG

SÄULE 2

LERN-
UMWELTEN

SÄULE 3

BILDUNGS-
ENTSCHEIDUNGEN

SÄULE 4

MIGRATIONS-
HINTERGRUND

SÄULE 5

BILDUNGS-
RENDITEN

ETAPPE 8

BILDUNG IM ERWACHSENENALTER UND LEBENSLANGES LERNEN

ETAPPE 7

HOCHSCHULSTUDIUM UND ÜBERGANG IN DEN BERUF

ETAPPE 6

ÜBERGÄNGE IN DIE BERUFLICHE AUSBILDUNG UND IN DEN ARBEITSMARKT

ETAPPE 5

WEGE DURCH DIE SEK. II UND ÜBERGÄNGE IN DEN TERTIÄREN BEREICH ODER ARBEITSMARKT

ETAPPE 4

WEGE DURCH DIE SEK. I UND ÜBERGÄNGE IN DIE SEK. II

ETAPPE 3

GRUNDSCHULE UND ÜBERGANG IN DIE SEKUNDARSTUFE I

ETAPPE 2

KINDERGARTEN UND ÜBERGANG IN DIE GRUNDSCHULE

ETAPPE 1

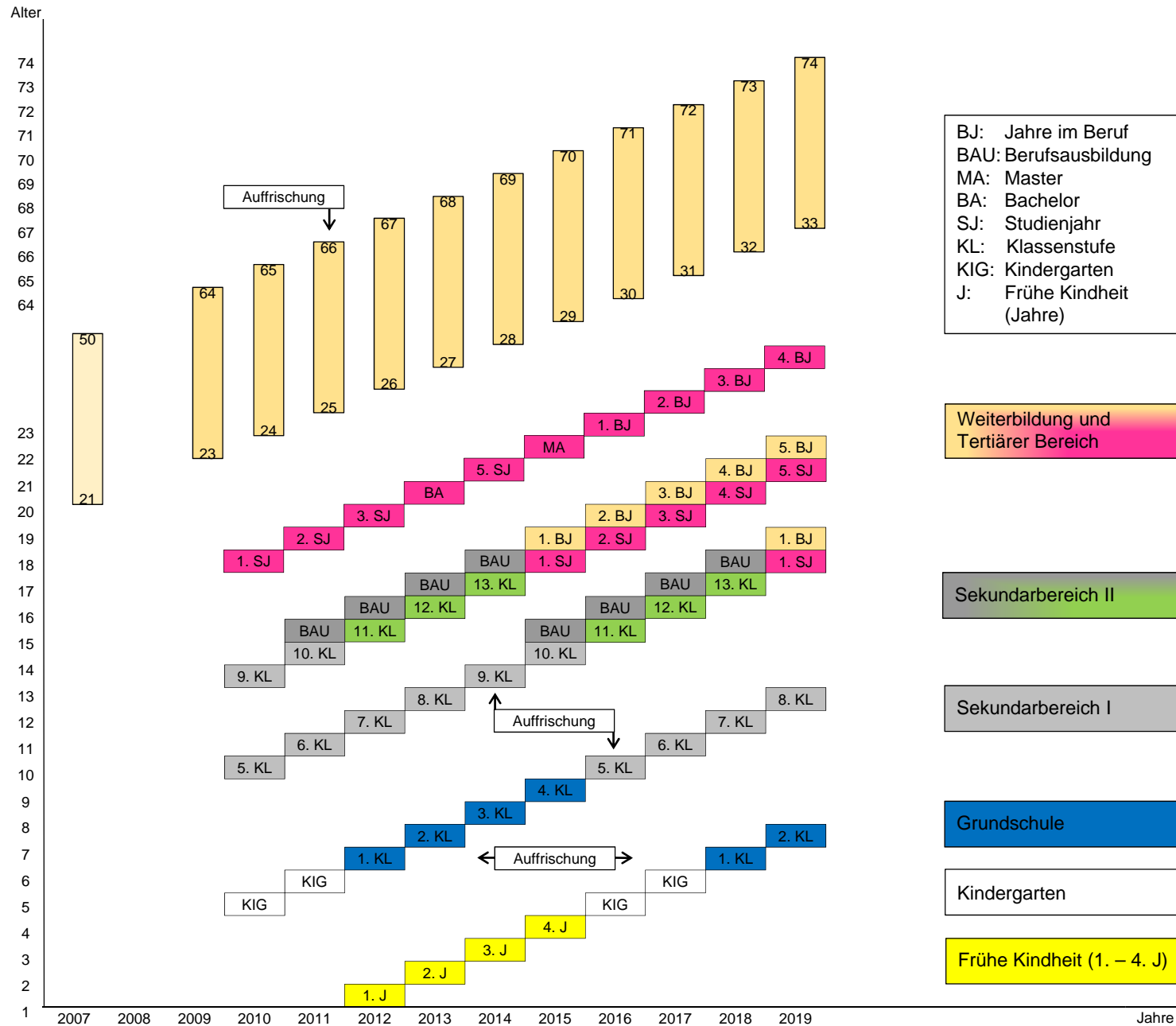
NEUGEBORENE UND FRÜHKINDLICHE INSTITUTIONELLE BETREUUNG

METHODENBEREICH

NUTZER-SERVICE, UMFRAGEMANAGEMENT, DATA-WAREHOUSE

Veröffentlichungstermine der NEPS-Daten

- **Sommer 2011:**
 - Erwachsene (SC6)
- **Sommer 2012:**
 - Kindergarten (SC2)
 - Schule 5. Klasse (SC3)
 - Schule 9. Klasse (SC4)
 - Erstsemester-Studenten (SC5)
- **Herbst 2014:**
 - Neugeborene (SC1)



Grenzen von „Flatfiles“

Beispiele:

- Fixe Struktur
- Alle nötigen Formate (Stata, SPSS, Mplus) müssen vorgehalten werden
- Grenzen bei der Archivierung

Planungen für das NEPS Data Warehouse

Zeitlich:

- Fileserver
- Relationale Datenbank
- Data Warehouse

Basis für neue Funktionalitäten

- Relationale Datenbank – atomisierte Speicherung der Daten.
- „Verlinkung“ der Daten über Primär- und Fremdschlüssel.
- In der Folge höhere Flexibilität im Umgang mit den Daten und
- Die Möglichkeit der Anwendung von Datenbankfunktionalitäten und Business Intelligence Tools.

Beispiel: Metadaten

Ablauf:

- Eingabe der Fragebögen.
- Übermittlung an Erhebungsinstitute.
- Eingang der Daten und Verknüpfung mit Metadaten (aus der Fragebogenerstellung).
- Verbindung von Daten und (einer Vielzahl von) Metadaten bei der Datenaufbereitung.

Alle Variablen liegen mit Metadaten angereichert vor.

Beispiel: Datenaufbereitung

Beispiele:

- Automatisierte Eingangskontrolle auf falsche oder unwahrscheinliche Werte.
- Änderungen (Versionierung) können an einzelnen Variablen vorgenommen werden.
- Automatisierung bei zu generierenden Variablen.
- Automatisierte Dokumentation aller Aufbereitungstätigkeiten.

Beispiel: Verschiedene Datenquellen

- Aktuelle Quellen bei NEPS: Die beiden Erhebungsinstitute.
- Mögliche weitere Quellen: Geodaten, Administrative Daten, weitere (Bildungs-) Studien.
- Gemeinsame Standards (z.B. DDI) und
- automatisierte Prozeduren beim Import der Daten (Plausibilitätsprüfung, Harmonisierung) erleichtern die gemeinsame Verwendung der Daten.

Beispiel: Individuelle Datenselektion I

- Vielfältige Analyseideen erschweren die Erstellung passender „Flatfiles“.
- Lösung: Individuelle Selektion der Variablen.
- Der Nutzer kann über ein Webinterface Variablen oder Variablensets selektieren und
- diese im gewünschten Format (Stata, SPSS) exportieren.

Beispiel: Individuelle Datenselektion II

- Als Unterstützung bei der Selektion können Datenbanktools genutzt werden.

Zum Beispiel:

- Die gespeicherten Metadaten.
- Häufigkeiten und Lagemaße.
- Weitergehende statistische Analysen.

Beispiel: Individuelle Datenselektion III

- Bei all diese Selektionen ist auf die Wahrung der Anonymität der Teilnehmer zu achten.

Lösungsansatz:

- Variablen werden mit „Sicherheitsstufen“ versehen (z.B. Skalen unproblematisch, Regionalangaben problematisch).
- Nur bestimmte Kombinationen sind bei der Selektion erlaubt.

Weitere Anwendungen

Beispiele:

- Langzeitarchivierung der Daten.
- Speicherung von Ergebnissen zur weiteren Verwendung durch andere Nutzer.
- Bei Nutzung von Business Intelligence Tools Verzicht auf weitere Statistiksoftware möglich.
- Automatisierung bei der Anonymisierung der Daten.

Weiteres Vorgehen

- Paralleler Aufbau von Fileserver und relationaler Datenbank.
- Tests der Funktionalitäten.
- Überprüfung des Nutzens für die Datenanalyse.
- Überprüfung im Hinblick auf Datenschutzaspekte.
- Zunächst Verwendung in internen Prozessen (Datenaufbereitung, Fragebogen- und Testerstellung).
- Freischaltung getesteter Funktionalitäten für die Scientific Community.