

# FDZ *Methodenreport*

Methodische Aspekte zu Arbeitsmarktdaten

Nr. 8/2008 (DE)

## **Datengenese zweier Datenkonzepte: MTG (Maßnahme-Teilnahme-Grunddatei) und ISAAK (Instrumente Aktiver Arbeitsmarktpolitik)**

**Eine Betrachtung ausgewählter Fälle am Beispiel der  
Förderung im Rahmen des ESF-BA-Programms**

*Astrid Engelhardt, Dirk Oberschachtsiek, Patrycja Scioch*

September 2008



**Bundesagentur für Arbeit**

## Inhaltsverzeichnis

<b>Abkürzungsverzeichnis</b> .....	<b>3</b>
<b>Abstract</b> .....	<b>4</b>
<b>1. Einleitung</b> .....	<b>5</b>
<b>2. Datengenerierung der ESF-BA-Statistik im Rahmen der Förderstatistik der BA (früher ST38) bis Ende 2004</b> .....	<b>6</b>
<b>3. Aufbau der Maßnahme-Teilnahme-Grunddatei (MTG)</b> .....	<b>10</b>
<b>4. Änderungen im Ziehungsverfahren der Quelldaten für das Monitoring</b> .....	<b>15</b>
<b>5. Neuaufbau der Individualdatenbank zur Maßnahmeteilnahmehistorik</b> .....	<b>18</b>
5.1 <i>Die DWH-Struktur der Datenbank ISAAK (Instrumente aktiver Arbeitsmarktpolitik)</i> .....	18
5.2 <i>Aufbau einer ISAAK-Auswerteschicht</i> .....	22
<b>6. Vergleich einer MTG- und ISAAK-Auswerteschicht</b> .....	<b>22</b>
6.1 <i>Strukturelle Unterschiede und zu erwartende empirische Differenzen</i> .....	23
6.2 <i>Analysedaten und Datenaufbereitung</i> .....	24
6.3 <i>Abgleich der Anzahl Kundennummern und Anzahl Sätze/Teilnahmen</i> .....	25
6.4 <i>Abgleich der Maßnahme- und Teilnahme-Merkmale</i> .....	26
6.4.1 <i>Ergebnisse bei der Förderung beruflicher Weiterbildung (ESF-FbW)</i> .....	27
6.4.2 <i>Ergebnisse bei der Maßnahmengattung Coaching</i> .....	31
6.4.3 <i>Erklärungsansatz</i> .....	33
<b>7. Zusammenfassung und Fazit</b> .....	<b>34</b>
<b>Anhang A</b> .....	<b>38</b>
<b>Anhang B</b> .....	<b>41</b>

---

**Abkürzungsverzeichnis**

AA	Agentur für Arbeit
ABM	Arbeitsbeschaffungsmaßnahme
BA	Bundesagentur für Arbeit
BewA	Bewerberangebot
BMAS	Bundesministerium für Arbeit und Soziales
coArb	computergestützte Arbeitsvermittlung der BA
COMPAS	computergestütztes Ausbildungsvermittlungs-System der BA
CORA	COSIMA Restrukturierung der Daten im Arbeitsamt
coSach	EDV-Unterstützung für die Sachbearbeitung der BA
COSIMA	Computersysteme im Arbeitsamt
coStat	computergestütztes Statistik-Informationssystem der BA
dezPDV	dezentrale Personendatenverwaltung der BA
DWH	Data Ware House
ESF	Europäischer Sozialfonds
ESF-BA-Programm	ESF-finanziertes und von der BA umgesetztes Programm des Bundesministeriums für Arbeit und Soziales
FbW	Förderung der beruflichen Weiterbildung
IAB	Institut für Arbeitsmarkt- und Berufsforschung der Bundesagentur für Arbeit
IEB	Integrierte Erwerbsbiografien
ISAAC	Instrumente Aktiver Arbeitsmarktpolitik
ITM	IT- und Informationsmanagement (Abteilung am IAB)
KS-Test	Kolmogorow-Smirnow-Test
MTG	Maßnahme-Teilnahme-Grunddatei
PSA	Personal-Service-Agentur
STADA	Statistische Datenbank der BA
VerBIS	Vermittlungs-, Beratungs- und Informationssystem der BA
zPDV	zentrale Personendatenverwaltung der BA

---

**Abstract**

Der vorliegende Beitrag dokumentiert die Datengenerierung der für wissenschaftliche Zwecke aufbereiteten Teilnahmemeldungen der Bundesagentur für Arbeit am Beispiel von Förderleistungen im Rahmen des ESF-Programms. 2004 und 2005 wurden systematische Änderungen in den datenliefernden Fachverfahren der Bundesagentur für Arbeit eingeführt. Diese Umstellung hatte eine neue Datenstruktur für die Erfassung von Maßnahmeteilnahmen durch das IAB zur Folge. Dies bedeutet, dass die Maßnahmeteilnahmen seit 2005 nicht mehr über die (alte) MTG-Struktur (MTG: Maßnahme-Teilnahme-Grunddatei) aufbereitet werden, sondern über eine neue Datenbasis namens ISAAK (Instrumente Aktiver Arbeitsmarktpolitik). Ein Vergleich zeigt, dass die Datenmengen bei ISAAK umfangreicher sind. Dies liegt zum einen an dem Ziehungsverfahren bei ISAAK, dass auch ältere Datenstände durch eine vollständige Neuladung aktualisiert. Zum anderen dürfte es damit zu tun haben, dass bisherige Regeln keine Verwendung mehr finden. In der MTG wurden Sätze nur dann als gültig angenommen, wenn auch ein Zugangssatz existiert - unabhängig davon ob Bestandssätze und Abgangssätze existieren. Die Ergebnisse lassen keinen materiellen Bruch in den Datenreihen zwischen ISAAK und MTG erwarten.

**Danksagung:**

Dieser Bericht ist im Rahmen der Begleitforschung zum ESF-BA-Programm durch das IAB entstanden. Die Begleitforschung wird im Auftrag des Bundesministeriums für Wirtschaft und Arbeit durchgeführt und aus Mitteln des Europäischen Sozialfonds sowie der Bundesagentur für Arbeit finanziert.

Wir danken ITM für die wertvolle Mithilfe und Beratung bei diesem Projekt, ohne die eine Umsetzung der vorliegenden Arbeit nicht möglich gewesen wäre. Ebenfalls danken möchten wir der Statistik-Abteilung der Bundesagentur für Arbeit sowie Axel Deeke und Thomas Kruppe für die wertvollen Hinweise und Anmerkungen, die zu dem Entstehen der Arbeit beigetragen haben.

---

## 1. Einleitung

Im Zeitraum 2000 bis 2007 gibt es zahlreiche mikroökonomische Evaluationsprogramme zur Untersuchung von Nettoeffekten der aktiven Arbeitsmarktpolitik, die erstmals im Rahmen einer systematisch aufgebauten Datenbasis(-bank) administrative Daten der Bundesagentur für Arbeit als Datengrundlage nutzen. Historisch sind dies in Deutschland die Evaluationsuntersuchungen zu Förderungen im Rahmen von ABM und zum ESF-BA-Programm (vgl. u.a. Deeke u.a., 2001 sowie Caliendo/Hujer/Thomsen, 2003). Mit dem Jahr 2004 zählen hierzu auch die Evaluationsanalysen von Förderungen im Rahmen der „Hartz-Gesetze“ (vgl. BMAS, 2006).

Voraussetzung für diese Evaluationsprojekte war eine strukturelle Verbesserung des Datenmaterials in Form allgemein nutzbarer Mikrodaten. Bis ca. 2000 konzentrierte sich die Aufbereitung der administrativen Daten nur auf einzelne Projekte oder auf die Verwertung aggregierter Eintritts-, Austritts- und Bestands- sowie Verbleibsstatistiken (vgl. Blaske/Plath/Nagel, 1992). Seither sind mehrere Anpassungen bei der Datenaufbereitung durch die Statistik der Bundesagentur für Arbeit und dem Institut für Arbeitsmarkt- und Berufsforschung (Abteilung: IT- und Informationsmanagement) vorgenommen worden. Das IAB hat auf dieser Grundlage eine Datenbasis konstruiert, die es zulässt Datenauszüge auf Individualebene zu Maßnahmeteilnahmen historisiert zu verknüpfen (vgl. Kruppe/Oertel, 2003).

2004 und 2005 wurden systematische Änderungen in den datenliefernden Fachverfahren der Bundesagentur für Arbeit eingeführt. Diese Umstellung hatte eine neue Datenstruktur für die Erfassung von Maßnahmeteilnahmen durch das IAB zur Folge. Aus der Perspektive des IAB bedeutet dies, dass die Maßnahmeteilnahmen seit 2005 nicht mehr über die (alte) MTG-Struktur (MTG: Maßnahme-Teilnahme-Grunddatei) aufbereitet werden, sondern über eine neue Datenbank namens ISAAK (Instrumente Aktiver Arbeitsmarktpolitik).

Bisher fehlt es an einer zusammenhängenden und zitierfähigen Dokumentation beider Datenbanken. Darüber hinaus ist bisher nur unzureichend beschrieben, welche quantitativen und qualitativen Auswirkungen die Systemumstellungen auf den statistischen Ausweis von Teilnahmemeldungen im Rahmen der IAB-Daten haben könnten.

Die vorliegende Arbeit greift diese Lücke auf und liefert eine Beschreibung der Datengenese zu IAB-Daten bei Teilnahmemeldungen. Exemplarisch wird auf Teilnahmemeldungen bei FbW (Förderung beruflicher Weiterbildung) und ESF-BA-Förderung (hier ESF-FbW und ESF-BA-Coaching – für eine Beschreibung der ESF-BA-Förderung siehe Deeke, 2004) zurückgegriffen. Dabei unterliegt die Erfassung der ESF-BA-Teilnahmemeldungen in mehrfacher Hinsicht Besonderheiten, auf die an geeigneter Stelle - und soweit zum Zeitpunkt der Bearbeitung dieser Dokumentation noch nachvollziehbar - hingewiesen wird. Darüber hinaus wird aufgezeigt, welche maßgeblichen Änderungen sich im Zeitverlauf seit 2000 bei der Genese der Teilnahmedaten

---

ergeben haben.<sup>1</sup> Zudem werden beide Datenbasen empirisch-vergleichend mit Teilnahmemeldungen desselben Gültigkeitszeitraums gegenübergestellt. Verwendet werden hierzu Teilnahmemeldungen zu den ESF-BA-geförderten Maßnahmearten FbW und Coaching. Ausgewählt wurden diese Maßnahmearten, da sie einen relativ großen Anteil an den Teilnahmen am ESF-BA-Programm haben und damit für die Evaluation im Rahmen der ESF-BA-Begleitforschung von besonderem Interesse sind.<sup>2</sup>

## **2. Datengenese der ESF-BA-Statistik im Rahmen der Förderstatistik der BA (früher ST38) bis Ende 2004**

Ausgangspunkt der Maßnahmeteilnahmedaten sind die in den Arbeitsagenturen erfassten Meldungen zu Teilnahmen von Personen an den verschiedenen Maßnahmen und Leistungen der aktiven Arbeitsmarktpolitik. Diese werden zunächst zur Verwaltung der Teilnahmen an arbeitsmarktpolitischen Maßnahmen verwendet. Erst in zweiter Linie werden aus den Meldungen aggregierte Statistiken erzeugt, die Grundlage des Monitoring der aktiven Arbeitsmarktpolitik sind. Die Genese dieser Geschäftsstatistiken bis Ende 2004 (hier am Beispiel der ESF-BA-Geschäftsstatistik ST38)<sup>3</sup> erfolgt in drei Schritten, die der Abbildung 1 entnommen werden können. Im Einzelnen lässt sich die Datengenese wie folgt skizzieren:

### a) Erfassung von Kundendaten

Jede Neuerfassung einer Person in einer Arbeitsagentur führt zunächst zu der Aufnahme eines Bewerberprofils (BewA). Hierbei werden Informationen zur Qualifikation und zur Person erhoben, die über das Fachverfahren coArb (computergestützte Arbeitsvermittlung) erfasst werden.<sup>4</sup> Neben Informationen zur Qualifikation und zur Haushaltssituation (u.a. Familienstand, Zahl der Kinder unter 16 Jahre) umfasst dies auch Personenmerkmale (Name, Anschrift, Geburtsdatum, Geschlecht, usw.).

---

<sup>1</sup> Die Vollständigkeit und die Exaktheit der Beschreibung der IT-Verfahren der Bundesagentur für Arbeit sowie die Dokumentation der Datengenese zu den Verfahren MTG und ISAAK ist ohne Gewähr. Im Laufe der Recherchen konnten keine zusammenhängenden Dokumentationen beider IT-Verfahren sowie zur Architektur der Maßnahme-Teilnahme-Grunddatei gefunden werden.

<sup>2</sup> Im Nachfolgenden werden die Begriffe Datenbank und Datenbasis synonym verwendet.

<sup>3</sup> Zu berücksichtigen ist hierbei, dass für die ST38 Teilnahmemeldungen direkt über das Fachverfahren CoSach erfasst wurden. Üblicherweise wurden Teilnahmen an einer spezifischen Maßnahme auf Individuenebene erst über das Verknüpfen von BewA-Merkmalen zur Teilnahme und CoSach-Maßnahmemerkmalen möglich. Dies erfolgte über das Verfahren CoStat (siehe unten).

<sup>4</sup> Diese Daten sind Ausgangspunkt der Bewerberangebotsdatei der Arbeitslosen und Arbeitsuchenden. Eine weitere potentielle Quelle der Bewerberangebotsdatei war das Verfahren Compas (computergestütztes Ausbildungsvermittlungs-System), was hier nicht weiter erläutert wird.

---

Unter anderem dienen diese Daten zur Verwaltung der Kundenprofile, die seit 1996 separat verarbeitet (Stammdaten) und für die Verknüpfung verschiedener Datenquellen auf Individualebene genutzt werden. Seither haben sich bei der Kundenerfassung einige Systemwechsel ergeben, die sich im wesentlichen auf die Zugriffsmöglichkeiten und die Verknüpfbarkeit der erfassten und verwalteten Personendaten beziehen.<sup>5</sup>

Zwischen 1996 bis Mitte 2003 existiert die Personendatenverwaltung nur über eine lokal operierende Anwendung, die eine Verwaltung des Kundenbestandes lediglich auf Arbeitsagenturebene ermöglicht. Zwar wurde bereits bis 2002 parallel eine zentrale Datenbank aufgebaut (CORA-Kundendatenbank), diese steht jedoch für zentrale Datenabgleiche erst seit 2004 vollständig zur Verfügung (zentrale Kundendatenverwaltung, zPDV).<sup>6</sup> Diese Systemanpassung hat zur Folge, dass für eine Person nunmehr mit höherer Sicherheit eindeutige Identifikatoren existieren.<sup>7</sup>

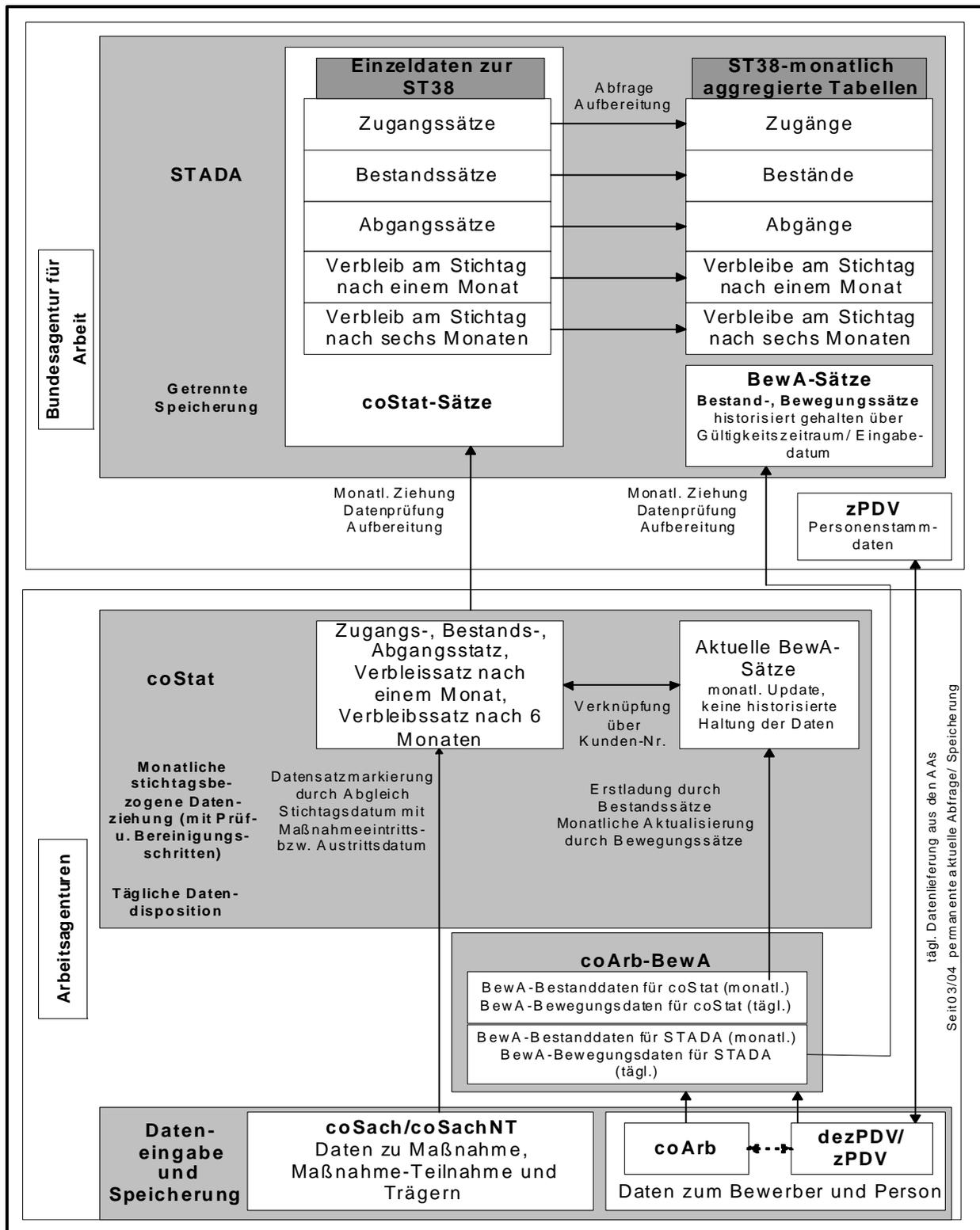
---

<sup>5</sup> Zeitlich vor 2000 liegende Umstellungen der IT-Verfahren werden hier nicht betrachtet. Größere Umstellungen gab es im Wechsel 2001/2002 durch die Implementation eines neuen Erfassungssystems.

<sup>6</sup> Mit dem neuen Verfahren (zPDV) besitzt nun jede Agentur über die zentrale Personendatenverwaltung (zPDV) einen aktuellen Zugriff auf die zentrale Personendatenbank der BA.

<sup>7</sup> Bis zu der Systemumstellung, konnte es vorkommen, dass eine Person – insbesondere durch Umzug – mit unterschiedlichen Identifikatoren geführt wurde.

---

Abbildung 1<sup>8</sup>: Datengenerese der Geschäftsstatistik ST38 bis Ende 2004

<sup>8</sup> In der Abbildung nicht aufgeführt ist die Tatsache, dass über coSach/coSachNT separat auch Personenmerkmale, wie Alter, Geschlecht u.ä. erfasst wurden.

#### b) Erfassung von Maßnahmeteilnahmen

Die Daten zu Maßnahmen, den Maßnahmeteilnahmen und zu den Trägern werden über ein zweites Fachverfahren erhoben. Dies erfolgt über das IT-Fachverfahren coSach (computergestützte Sachbearbeitung) und seit Dezember 2004 durch die erweiterte Version coSachNT (computerunterstützte Sachbearbeitung von Leistungen zur Beschäftigungsförderung).<sup>9</sup> Das Besondere an diesem Verfahren (coSach und coSachNT) ist, dass es eine Schnittstelle zu den Personendaten enthält, so dass eine Verwaltung von Teilnehmern an einer Maßnahme möglich ist. Um diese Teilnahmen eindeutig identifizieren zu können, verwaltet dieses Fachverfahren jede Maßnahmeteilnahme eines Kunden über einen eigenen Identifikator (Maßnahmeteilnahmenummer).

#### c) Statistische Aufbereitung von Kundeninformationen und Teilnahmedaten

Die beiden oben genannten Fachverfahren (coArb und coSach) dienen hauptsächlich der Verwaltung von Personen- und Teilnahmedaten und lassen sich für statistische Zwecke, nur bedingt nutzen. Für diese Zwecke existierte ein weiteres Fachverfahren (coStat: computergestützte Statistik), das den Agenturen (bis 30.03.2005) lokal zur Verfügung stand. Durch dieses IT-gestützte Informationssystem wurde auch der Austausch der Daten für statistische Zwecke mit der zentralen Datenbank der Bundesagentur für Arbeit organisiert. Dieser erfolgte einmal im Monat, zu einem bestimmten Stichtag. Die Versorgung des lokalen Statistik-Fachverfahrens erfolgte täglich.<sup>10</sup>

Um für die weitere statistische Erfassung Maßnahmeteilnahme-Zugänge und -Abgänge voneinander trennen zu können, waren bei der Datenziehung bestimmte Bereinigungs-schritte erforderlich. Dabei wurden die Teilnahmesätze als Zugangs-, Bestands-, Abgangs- und Verbleibssätze (sog. Z-, B-, A-, V-Sätze) markiert, wobei diese aufgrund einfacher Prüfroutinen durch einen Datumsabgleich zwischen Ziehungsstichtag und Maßnahmeteilnahmeeintrittsdatum bzw. -austrittsdatum erzeugt wurden.<sup>11</sup> Zusätzlich wurden weitere Prüf- und Bereinigungs-schritte vorgenommen, die gewährleisten, dass nur konsistente und fehlerfreie Meldungen für die zentrale Statistik berücksichtigt werden. Außerdem wurden

---

<sup>9</sup> Bei der Datengenese der Teilnahmen zur ESF-BA-Förderung (ST38-Teilnahmedaten) ist zu berücksichtigen, dass hier Teilnahmen (betrifft die Zugangs-, Bestands und Abgangssätze) direkt über das Fachverfahren coSach identifiziert werden konnten (und nur an coStat übergeben wurden). Üblicherweise wurde die Maßnahmeteilnahme (Teilnahme eines Kunden an einer spezifischen Maßnahme) erst über das Verfahren coStat durch eine Verknüpfung der CoSach-Maßnahmeninformationen und der Kunden-Teilnahmeinformationen möglich – also erst dort identifizierbar.

<sup>10</sup> Die tägliche Versorgung durch sogenannte Bewegungssätze ist Ursache für mögliche Mehrfachsätze zu einer Teilnahmemeldung.

<sup>11</sup> Eine genauere Beschreibung zum Vorgehen findet sich im Anhang (Anmerkung A1)

die Maßnahmeteilnahme-Datensätze mit ausgewählten Merkmalen aus der Bewerberangebotsdatei und technischen Merkmalen angereichert.

Mittels CoStat wurden schließlich die Daten in Auszügen zentral zusammengezogen und für die Aufbereitung der Statistik bereitgestellt.<sup>12</sup> Diese in STADA verwalteten Sätze waren Grundlage der BA-Geschäftsstatistiken, die in Form von monatlich aggregierten Tabellen zu Zugängen, Beständen, Abgängen und Verbleiben bei Maßnahmeteilnahmen und der Arbeitsuchendenstatistik erstellt werden.<sup>13</sup>

### 3. Aufbau der Maßnahme-Teilnahme-Grunddatei (MTG)<sup>14</sup>

Aus Forschungsperspektive sind die für die Geschäftsstatistik aufbereiteten Daten aus mehrfacher Sicht unzureichend:

- 1) Das STADA Konzept war stichtagsbezogen und dient lediglich für einfache – vor allem aggregierte - statistische Auszählungen.
- 2) Dies bedeutet, dass die Zugänge in eine Maßnahme dem Monat zugerechnet sind, in dem sie gezogen und damit statistisch erfasst werden. Eine verzögerte Eingabe der Teilnahmedaten führt dadurch zu zeitlich nicht korrekt ausgewiesenen Teilnehmerzahlen.
- 3) Ebenso folgt aus diesem Stichtagskonzept und dem damit verbundenen Ziehungsverfahren, dass ggf. bei verspätet erfassten Teilnahmemeldungen zeitlich ungültige - nämlich aktuellere - BewA-Merkmale zugespielt wurden. Die Teilnahmemeldungen und die BewA-Merkmale in einem Datensatz können somit unterschiedliche Gültigkeitszeiträume aufweisen.
- 4) Darüber hinaus wurden von coStat für jede Teilnahme unverbundene Einzeldatensätze zum Zugang, Bestand, Abgang und Verbleib (1. und 6. Monat nach Austritt) erzeugt.

---

<sup>12</sup> Auf die genaue Verarbeitung und Organisation der BewA-Sätze in coStat und STADA wird an dieser Stelle nicht weiter eingegangen (siehe hierzu Kruppe/Oertel, 2003). Hingewiesen sei aber darauf, dass es zwei unterschiedliche BewA-Satzarten gab.

<sup>13</sup> Teilnahmesätze, für die eine dezentrale Zuspiegelung von BewA-Merkmalen nicht möglich war, weil in der Arbeitsagentur kein BewA-Datensatz gefunden wurde, wurden nicht berücksichtigt. Das BewA wurde am Wohnort-Arbeitsamt geführt. Fand die Teilnahme an einer Maßnahme in einem anderen Arbeitsamt-Bezirk statt, wurde die Teilnahme in der dortigen coSach-Datenbank registriert (Ausnahme: ESF-BA-Geförderte müssen am Wohnort-AA geführt werden).

<sup>14</sup> Die folgenden Beschreibungen erheben nicht den Anspruch auf Vollständigkeit. Der Aufbau der MTG-Datenbank und ihrer Auswerteschichten mit den einzelnen dabei durchgeführten Prüf- und Bereinigungsschritten ist zu komplex, um sie hier im Einzelnen zu beschreiben. Die folgende Darstellung ist daher bewusst reduziert.

---

Durch diese Eigenschaften sind die STADA-Daten für wissenschaftliche Zwecke – insbesondere Mikroanalysen – nur von eingeschränktem Nutzen, da sie keine widerspruchsfreien, zeitlich konsistenten Merkmalskombinationen auf individueller Ebene ermöglichen. Die über STADA vorliegende Datenstruktur erfordert daher ein für die Forschung abweichendes Vorgehen der Erschließung und Aufbereitung der Individualdatensätze aus den operativen IT-Verfahren der BA, bei dem die „Verwaltungsdaten“ zu „Forschungsdaten“ transformiert werden (Kruppe/Oertel, 2003).

Mit der MTG (Maßnahme-Teilnahme-Grunddatei) wurde versucht, die Defizite des Datenausgangsmaterials zu kompensieren. Ziel des IABs war es, eine Individualdatenbank zu erstellen, welche den Anforderungen – insbesondere zum Zwecke der Evaluation – gerecht wird und es ermöglicht, Maßnahme-, Erwerbs- und Arbeitslosigkeitsverläufe auf Personenebene zu verknüpfen und historisiert zu verfolgen.<sup>15</sup> Dabei wurden mehrere Ziele verfolgt:

- 1) die Verknüpfung der nach Zugang, Bestand und Austritt getrennten Datensätze der individuellen Teilnahmen,
- 2) die eindeutige Berücksichtigung der Eintritts- und Austrittsdaten bei der zeitlichen Zuordnung und
- 3) die Beachtung der „zeitkritischen“, d. h. im Förderverlauf veränderlichen Merkmale,
- 4) es sollte eine entsprechend auf die Verbleibsstichtage bezogene Verbleibsermittlung und schließlich eine über die jeweilige Teilnahme hinausgehende „Historisierung“ auf Personenebene erfolgen.

Mit der MTG wurde ab 2001 ein entsprechendes Datenbank-Konzept erarbeitet, das diese Anforderungen erfüllen sollte. Die MTG (vgl. Abbildung 2) erhielt hierzu ihre Maßnahme- und Teilnahmedaten aus allen seit Beginn 2000 bis Ende 2004 monatlich erstellten Einzeldaten der BA-Geschäftsstatistiken, die über STADA zur Verfügung standen. Die coStat-Sätze wurden im weiteren Prozess einmal im Monat in die MTG-Ladeschicht geladen, wobei nur die coSach/coSachNT-Informationen übernommen wurden. Die BewA-Merkmale aus den Geschäftsstatistiken/STADA blieben aus den oben genannten Gründen (zeitkritische Attribute) unberücksichtigt. Stattdessen wurden die Informationen aus den BewA-Bestandssätzen verwendet, die ebenfalls über STADA verfügbar waren und historisiert vorlagen.

Zusätzlich zu den Prüfroutinen bei der Datenziehung für STADA wurden weitere Regeln und Bereinigungen beim Laden der MTG angewendet. Hierbei sind die Dimensionen, die Datumsfelder

---

<sup>15</sup> Ergebnis dieser Bemühungen waren zunächst zwei separate Datenbanken mit ähnlich strukturierten Datensätzen, die parallel und exemplarisch zur Erstellung der MTG aufgebaut wurden: a) die ESF-BA-Individualdatenbank und b) die ABM-Teilnehmerdatenbank. Beide Datenbasen wurden in die MTG integriert.

und die numerischen Werte überprüft worden. Die Daten lagen damit, ähnlich wie in STADA, als Zugangs-, Bestands-, Abgangs- und Verbleibssätze (nach einem bzw. nach sechs Monaten) vor. Maßgeblich für das weitere Zusammenspielen der Datentabellen zu Auswerteschichten<sup>16</sup> waren dann drei zentrale Entitäten:

#### 1. Entität Maßnahme

Die Ausgangsdaten der Entität Maßnahme waren die Maßnahmedaten der Geschäftsstatistiken. Dies waren (zunächst) die Daten zu Maßnahmen der Förderung beruflicher Weiterbildung (FbW), darunter auch ESF-BA-geförderte, und Arbeitsbeschaffungsmaßnahmen (ABM). Eine Maßnahme war hierbei eindeutig durch ihre Maßnahmenummer festgelegt. Maßnahmen für die keine Maßnahmedaten sondern nur Teilnahmedaten vorlagen blieben bei diesem Vorgehen jedoch unberücksichtigt. In der MTG wurden diese künstlich durch die maßnahmespezifischen Merkmale der Teilnehmerdaten erzeugt (z. B. Maßnahmenummer).<sup>17</sup>

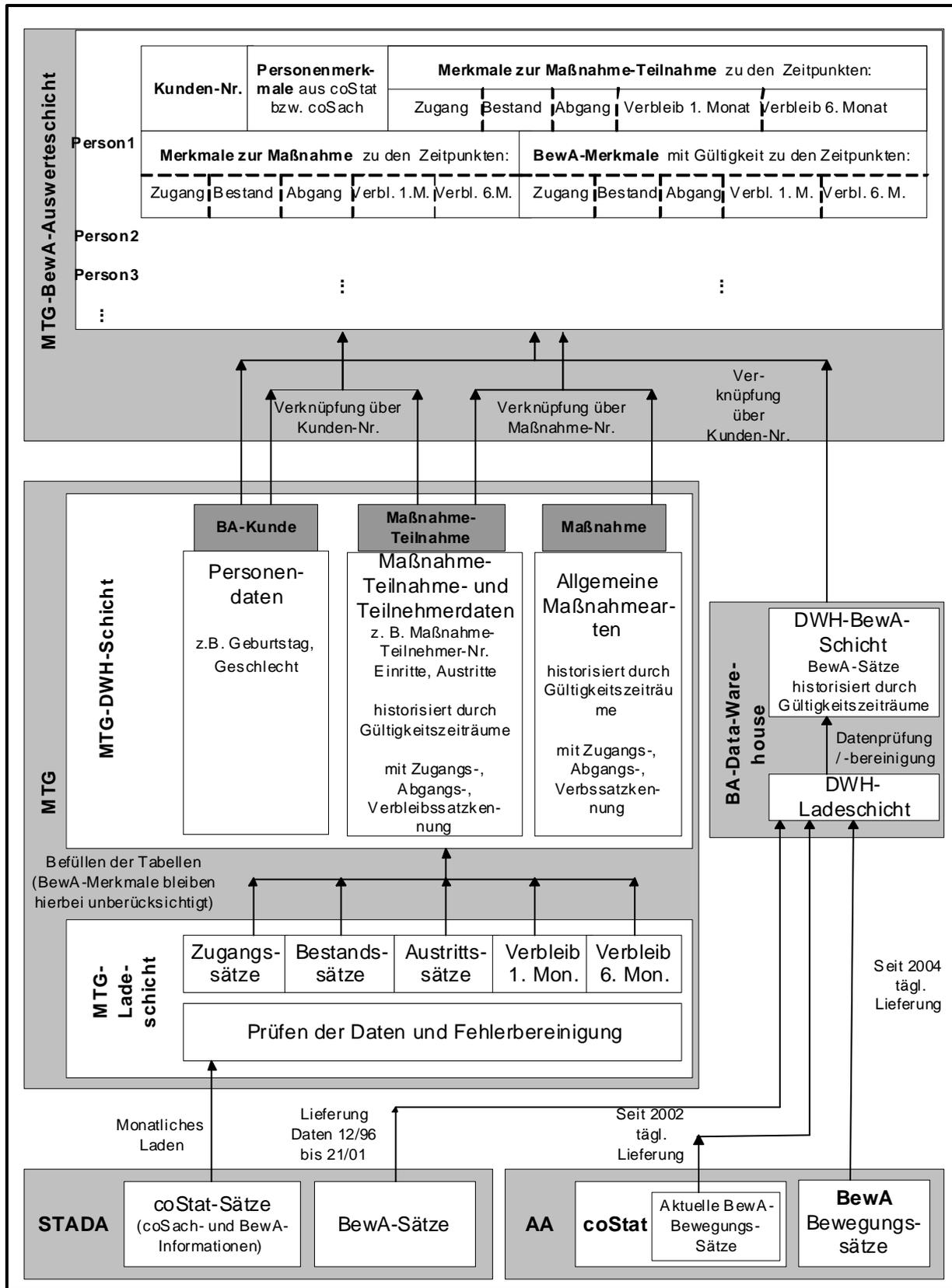
---

<sup>16</sup> Auswerteschichten sind Teilmengen der MTG, die für spezifische Datenauswertungen oder Anwendungen über die Datenstruktur der MTG erzeugt werden.

<sup>17</sup> Dies betrifft neben der Freien Förderung und Deutschlehrgängen auch alle ESF-Maßnahmen, welche nicht in Zusammenhang mit einer FbW-Maßnahme stattfinden.

---

**Abbildung 2: Aufbau der Maßnahme-Teilnahme-Grunddatei und einer MTG-BewA-Auswerteschicht**



## 2. Entität BA-Kunde

Diese Entität basiert auf den über die Personendatenverwaltung erfassten Kundendaten, die mit der zentralen Personendatenbank CORA (vgl. Abschnitt 1) zur Verfügung stehen. Der Anfangsbestand (1996 bis 2001) dieser Entität wurde über eine Bestandsladung zum Oktober 2001 erzeugt und wird durch monatliche Neuladungen aktualisiert. Zu den hier gespeicherten Personendaten zählten z. B. das Geschlecht und das Geburtsdatum der Personen. Zur Identifikation dienten die Attribute BewA-Eingabenummer und BA-Kundenummer.

## 3. Entität Teilnahme des BA-Kunden

Die Tabellen der Entität Maßnahme-Teilnahme füllten sich mit den Teilnehmerdaten aller Maßnahmearten aus coStat. Die Teilnehmernummer identifiziert hierbei eindeutig die Teilnahme an einer Maßnahme. Die Merkmale zur Maßnahmeteilnahme, wie z. B. das Eintritts- und Austrittsdatum, wurden über die Gültigkeitszeiträume historisiert gespeichert. Auch die Kennzeichnung als Zugangs-, Bestands-, Abgangs- oder Verbleibssatz war hier gesichert.

Die Maßnahmedaten ließen sich dann über spezifische Identifikatoren mit den Teilnahmedaten und mit den Kundenmerkmalen verknüpfen. Auf diese Weise wurde erreicht, dass für eine Person alle zuordenbaren Teilnahmen beobachtbar waren und dass die Personenmerkmale (BewA-Merkmale) weitestgehend zeitlich adäquat zum Zeitpunkt der Maßnahmeteilnahme vorlagen.

Konkret enthielt in einer MTG-Auswerteschicht jeder Satz eines Teilnehmers die Kundenummer, Personenmerkmale und die Maßnahme- und Teilnahmemerkmale jedes coStat-Satzes, mit einer Markierung, die den Quelldatensatz (Zugang, Abgang, Verbleib 1. Monat oder Verbleib 6. Monat) vermerkte. Da die MTG-Auswerteschicht durch dieses Vorgehen für Auswertungen zu wenig personenbezogene und erwerbsbiografische Merkmale aufwies, wurde sie für Auswertezwecke mit Merkmalen aus der Bewerberangebotsdatei angereichert (zusammen mit den MTG-Sätzen ergibt sich dann eine sogenannte MTG-BewA-Auswerteschicht). Die hierbei verwendeten BewA-Sätze wurden direkt aus der DWH-Schicht des BA-Data-Warehouse geladen (vgl. Abbildung 2). An den entstandenen Satz wurden somit zu allen in dem Satz ausgewiesenen Eintritts- und Austrittsdaten (Maßnahme-Teilnahme-Daten) die zu den jeweiligen Zeitpunkten gültigen BewA-Merkmale herangespielt. Auch die BewA-Merkmale erhielten hierbei eine Kennung, die über das jeweilige Datum und den Quelldatensatz Auskunft gab. Eine genauere Erläuterung des BA-Data-Warehouse (BA-DWH) sowie eine Beschreibung des Aufbaus der DWH-BewA-Schicht erfolgt im Anhang (Anmerkung A2 und A3). Neue Sätze wurden jeweils neu an die MTG herangespielt.

---

#### 4. Änderungen im Ziehungsverfahren der Quelldaten für das Monitoring

Die der MTG zugrundeliegende Datenarchitektur hat sich seit 2005 grundlegend verändert<sup>18</sup>:

##### 1. Abschalten von coStat und STADA

Das Abschalten von coStat und STADA hat zur Folge, dass die BewA-Merkmale der Personen nicht mehr dezentral „vor Ort“ mit den Merkmalen der entsprechenden Maßnahmeteilnahmen aus coSachNT verknüpft werden (z.B. bei ESF-FBW-Teilnahmen wurden die BewA-Merkmale aus dem DWH entnommen: STADA-BewA-Sätze). Darüber hinaus entfällt die Markierung der Sätze als Zugangs-, Bestands-, Austritts- und Verbleibssätze durch coStat bzw. coSach (ESF-BA-Besonderheit).<sup>19</sup>

Bisher erfolgte die Anschließung der BewA-Merkmale an die coSach-Merkmale einer Person dezentral in coStat auf der Ebene der AA. Im DWH liegen jetzt die BewA-Informationen aller Agenturen zentral vor, so dass auch bei einer Abweichung von Wohnort- und Maßnahme-Agentur diese Informationen hinzugespielt werden können.<sup>20</sup> Dies hat einerseits eine Erhöhung der Fallzahlen bei Neumeldungen in den Statistiken zur Folge. Andererseits fallen sehr stark verspätete Teilnahme-Neumeldungen aus den Statistiken heraus. Gleichzeitig werden Maßnahmeteilnahmen mit einer Zeitverzögerung von bis zu 3 Monaten für den relevanten Monat bestandswirksam. Im Altverfahren wurden verspätete Maßnahmeteilnahmen für den Monat des Dateneingangs gezählt.

##### 2. Einführung des Fachverfahrens coSachNT

Mit der Einführung von coSachNT existieren Schnittstellen, die einen direkten Datenaustausch zwischen einzelnen Fachverfahren ermöglichen. Dies sorgt für eine höhere Qualität in den verarbeiteten Daten. Z.B. gibt es mit coSachNT eine direkte Anbindung der Maßnahmeteilnahme zur zPDV, so dass nur Eingaben von Kundennummern zulässig sind, die bereits über die zentrale Personendatenbank (zPDV) erfasst sind. Außerdem sind in coSachNT weitere Prüfroutinen implementiert, die z.B. einen Abgleich zulässiger Förderzeiten ermöglichen und es besteht ein

---

<sup>18</sup> Die Systemumstellung von coSach auf coSachNT wurde bereits 2003 gestartet. Ein vollständiges „Abschalten“ der coStat-Sätze sowie die Beendigung der Versorgung der STADA-Datenbank erfolgten erst Ende 2004. Folgendes sei zudem angemerkt: Bei den Ausführungen sollte berücksichtigt werden, dass die Datengenese und die Umstellungen mit 2005 überwiegend anhand der ESF-FBW und der FBW-Förderung nachvollzogen wurden. Soweit recherchierbar wird auf mögliche Besonderheiten bei der Erfassung und der Datenaufbereitung dieser Teilnahmenmeldungen hingewiesen.

<sup>19</sup> Für das Monitoring werden die neuen Abfragen aus dem DWH durchgeführt, bei denen eine künstliche Markierung der Sätze als Zugangs-, Bestands-, Austritts- und Verbleibssatz (1. und 6. Monat) erfolgt. Im DWH werden keine Sätze im Z, B, A, V- Format verwaltet.

<sup>20</sup> Dies betrifft insbesondere die Maßnahmart „Qualifizierung während Kurzarbeit“. Die Teilnehmer an diesen Maßnahmen sind in der Wohnortagentur arbeitsuchend gemeldet, führen die Qualifizierungsmaßnahme aber ggf. in einem Betrieb durch, für den eine andere Arbeitsagentur zuständig ist.

bundesweiter Zugriff auf die (jetzt vollständig) in coSachNT erfassten Maßnahme- und Maßnahmeteilnahmedaten.

### 3. Ablösung von coArb durch VerBIS

Größere Änderungen haben sich auf der Seite der Personendaten ergeben. Die Daten zur Person bzw. zum Bewerber für die Bewerberangebotsdatei wurden bisher im Fachverfahren coArb eingegeben und zunächst lokal gespeichert. Dieses Fachverfahren ist in den Agenturen zwischen Juli 2005 bis Juni 2006 sukzessiv von dem Verfahren VerBIS (Vermittlungs-, Beratungs- und Informationssystem) abgelöst worden. Seit Mitte des Jahres 2006 erfolgen die Neueingaben und Aktualisierungen der Bewerberangebote (und Stellenangebote) ausschließlich in VerBIS. Die Kundenstamm-Daten werden aber weiterhin über die zPDV versorgt.

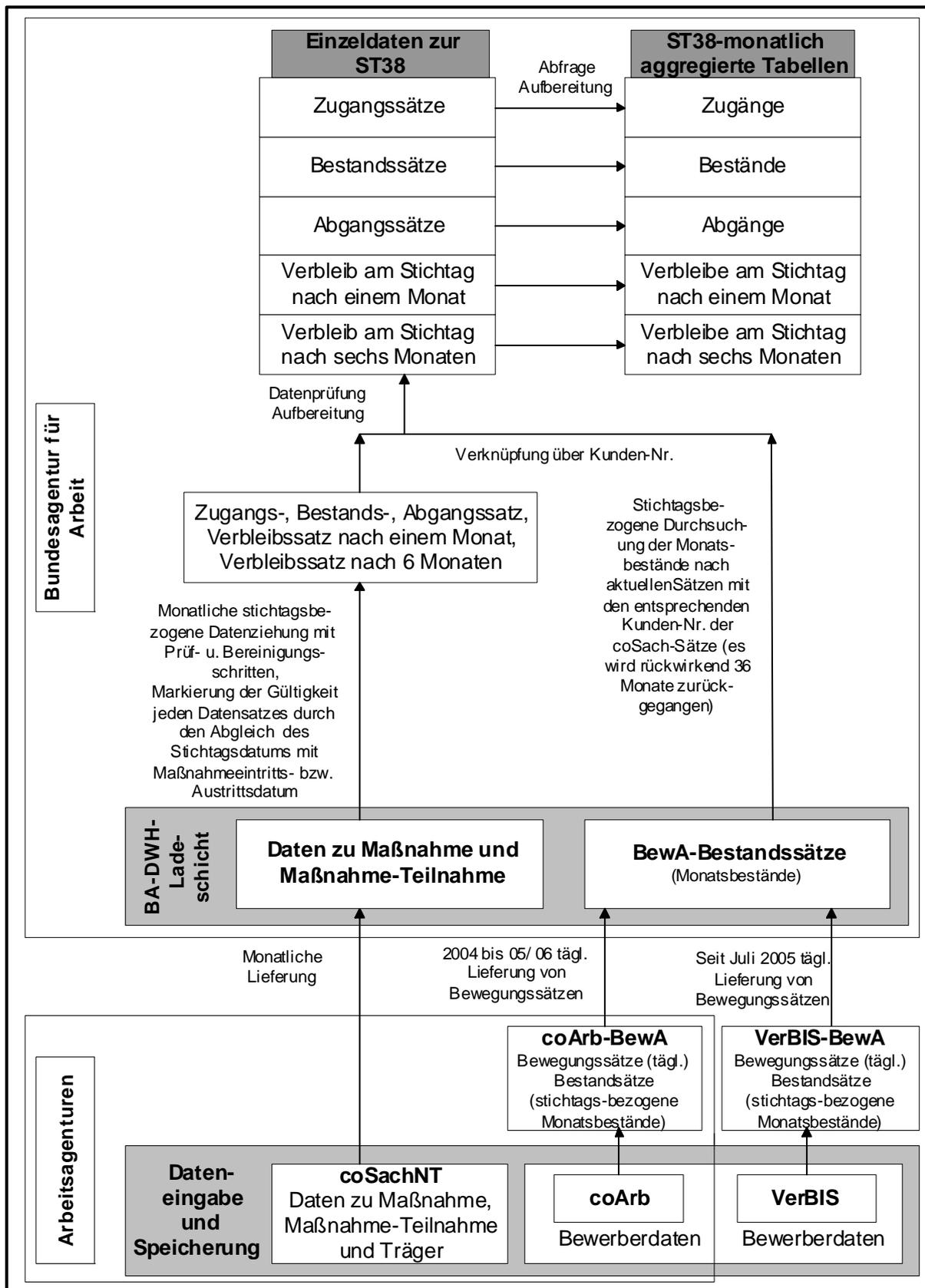
Anders als das dezentrale Fachverfahren coArb der einzelnen Agenturen, ist VerBIS ein Verfahren, welches die Daten an einem Ort zentral verwaltet und auf das jede Agentur zugreifen kann. Aus VerBIS werden täglich die dort gespeicherten BewA-Bewegungssätze, analog zu coArb, an das DWH geliefert.<sup>21</sup> Bei der Verknüpfung der coSachNT-Informationen mit den BewA-Sätzen werden am Stichtag die Monatsbestände der Ladeschicht des DWHs nach BewA-Sätzen mit den entsprechenden Kundennummern durchsucht. Es wird dabei im Datenstand bis zu 36 Monate (vom Stichtag ab) zurückgegangen, um für jede der Personen einen Ladesatz zu finden. Die weitere Aufbereitung der Daten zu der monatlichen ESF-Förderstatistik erfolgt analog der coStat/STADA-Logik (vgl. Kapitel 1)

---

<sup>21</sup> Auf die Probleme bei der Verarbeitung der Daten aus VerBIS während des Parallelbetriebes mit coArb und nach dessen Abschalten wird hier nicht näher eingegangen. Ebenso wird die Problematik der Datenbrüche in den BewA-Daten aufgrund der Umstellung der Fachverfahren nicht thematisiert.

---

Abbildung 3: Datengenese der ESF-Förderstatistik (früher ST38) seit Anfang 2005



## 5. Neuaufbau der Individualdatenbank zur Maßnahmeteilnehmehistorik

### 5.1 Die DWH-Struktur der Datenbank ISAAK (Instrumente aktiver Arbeitsmarktpolitik)

Durch den Wegfall des STADA-Datenbanksystems ist die Versorgung der MTG mit Daten aus coStat im Prinzip nicht mehr möglich. Gleichzeitig sollten einige Qualitätsverbesserungen bei der Erstellung einer Datenbasis zu historisierten Maßnahmeteilnahmen erreicht werden. Beide Aspekte führten 2004/2005 dazu, dass im IAB eine neue Datenbank mit dem Namen ISAAK (Instrumente aktiver Arbeitsmarktpolitik) aufgebaut wurde. Diese enthält Maßnahmeteilnahmen von Personen an der aktiven Arbeitsmarktpolitik der BA und löst die MTG ab.

Der Aufbau von ISAAK, die Datenversorgung wie auch die Aufbereitungs- und Bereinigungs-schritte beim Laden der Daten haben sich gegenüber der MTG grundlegend verändert (vgl. Abbildung 4<sup>22</sup>). Die Datenbank wird alle drei Monate komplett neu geladen. Die Datenversorgung von ISAAK erfolgt nunmehr ausschließlich aus dem DWH der Bundesagentur für Arbeit. Dabei werden die Maßnahme-, Bewerberdaten und die weiteren personenspezifischen Merkmale aus der BA-DWH-Schicht (BewA und zPDV) entnommen und unterliegen damit bereits weiteren Prüfroutinen und Bereinigungs-schritten. Die Teilnahmedaten dagegen kommen direkt aus der BA-DWH-Ladeschicht<sup>23</sup> und liegen nicht mehr für unterschiedliche Zeitpunkte vor (Zugang, Bestand, Abgang und Verbleib).<sup>24</sup> Im Prinzip wird mit ISAAK eine eigene DWH-Struktur aufgebaut, bei der mehrere Entitäten existieren, die wechselseitig für Prüfroutinen genutzt werden.<sup>25</sup>

#### 1. Die ISAAK-Person

Der Aufbau der ISAAK-Person erfolgt zunächst über eine vollständige Ladung des Bestandes an Personendaten aus der DWH-Schicht (vgl. Abbildung 4). Personen aus der zPDV mit jeweils einer Kundennummer erhalten anschließend einen eindeutigen personenspezifischen Identifikator. Personen mit Teilnahmesätzen die keine BA-Kundennummer aber eine BewA-Eingabenummer besitzen und über diese keine Kundennummer gefunden werden kann (dies

---

<sup>22</sup> Ebenso wie bei der MTG wird bei der Beschreibung des Aufbaus und des Inhaltes von ISAAK aufgrund ihrer Komplexität nicht auf jedes Detail eingegangen. Es wird auf die Inhalte eingegangen, die für die Erstellung einer ISAAK-Auswerteschicht relevant sind.

<sup>23</sup> Außer bei FbW und PSA werden alle maßnahmebezogenen Attribute aus den Teilnahmesätzen genommen. Die FbW-Maßnahme-, Bewerber-, und Personendaten sind dagegen aus der BA-DWH-Schicht, die einmal im Monat über die BA-Ladeschicht mittels diverser Prüfungs- und Bereinigungs-schritte versorgt wird.

<sup>24</sup> Je nach Zeitraum ergeben sich für die beiden Datenbanken jedoch unterschiedliche Datenquellen. Die Maßnahme- u. Maßnahmeteilnahmedaten von Januar 2000 bis Mitte August 2002 stammen aus der STADA-Datenbank und sind monatlich gelieferte coStat-Sätze. Die Daten von Mitte August 2002 bis jetzt sind monatliche Lieferungen direkt aus coSach/coSachNT. Die Bewerberdaten sind aus co-Arb bzw. VerBIS. Daten des Zeitraumes 12/96 bis 12/01 stammen aus dem STADA-Bestand an BewA-Bestands- und Bewegungssätzen, BewA-Bewegungssätze von 01/02 bis 12/03 lieferten die Agenturen dezentral täglich über coStat und seit Anfang 2004 senden sie die Sätze direkt an das DWH. Die Personendaten werden täglich aus der zPDV geliefert.

betrifft in der Regel nur Daten der Jahre 1997 bis 1999), bekommen hier auch einen personenspezifischen Identifikator zugeordnet.

Liegen in der Gruppe der Teilnehmer mit coSachNT-Meldungen bei einer Eingabenummer mehrere Kundennummern vor - oder umgekehrt - werden für die Identifikation einer eindeutigen ISAAK-Person zusätzliche DWH-Relationen berücksichtigt. Neben der Beziehung Kundennummer - BewA-Eingabenummer werden die entsprechenden Gültigkeitszeiträume der BewA-Sätze einbezogen. Abschließend wird zu jeder eindeutigen ISAAK-Person die entsprechende Personenhistorik auf Basis der BewA- und zPDV-Attribute erstellt.<sup>26</sup>

## 2. Die ISAAK-Maßnahmeteilnahme

Die ISAAK-Maßnahmeteilnahmen werden zunächst über eindeutig identifizierbare coSachNT-Meldungen den ISAAK-Personen zugeordnet. In den Ladetabellen von ISAAK können allerdings zu einer Teilnahme einer ISAAK-Person mehrere Sätze vorhanden sein, so dass Teilnahmen nicht eindeutig identifiziert werden können. Dies gilt vor allem für Daten aus dem Altverfahren coStat. Bezogen auf aktuellere Meldungen gilt dieses auch, wenn in relevanten Attributen Änderungen vorgenommen wurden, wodurch ggf. neue Sätze erzeugt werden können. In solchen Fällen wird aus den Mehrfachsätzen zu einer Teilnahme nur derjenige Satz behalten, der die Informationen am „besten“ wiedergibt (Kriterien hierfür sind u.a. die Ausprägungen des Eintritts- und Austrittsdatums). Um diesen Satz je Teilnahme zu identifizieren, wird die Entität Maßnahme-Teilnehmer-Biografie aufgebaut, in der die Daten in so genannte Teilnahmegruppen eingeteilt werden.

In diesen Teilnahmegruppen sind alle Sätze enthalten, die eine identische Kombination aus Kundennummer, Maßnahmennummer und Teilnehmernummer aufweisen. Bei Teilnahmen an Trainingsmaßnahmen wird diese Regel anders ausgefüllt. Hier wird nur über die Kundennummer und die Maßnahmennummer gruppiert. Nicht-Beste Sätze werden in einer eigenen Tabelle abgelegt, die den Namen Rucksack trägt.

Das hierbei verwendete Regelwerk stützt sich wie folgt auf das erfasste Eintritts-, Austritts- und Erstelldatum der Sätze:

Überschneiden sich bei allen Sätzen einer Teilnahmegruppe die Teilnahmezeiträume oder sind identisch, so wird der Satz mit dem aktuellsten Erstelldatum als der „beste“ angesehen. Dies geschieht unter der Voraussetzung, dass er nicht mehr als 365 Tage nach Teilnahmeende eingegeben wurde.

---

<sup>25</sup> Im Prinzip wurde dies bereits mit der MTG realisiert. Der Aufbau der Relationen wurde bei ISAAK jedoch spezifisch berücksichtigt und detaillierter ausgearbeitet.

<sup>26</sup> Technisch wird hierzu eine eigene Entität „ISAAK-Bewerber“ aufgebaut. Die Zuordnung zum BA-Kunden bzw. zur ISAAK-Person erfolgt über eine M:N-Beziehung.

---

Gibt es Sätze einer Teilnahmegruppe deren Teilnahmezeiträume überschneidungsfrei sind, so werden alle diese Sätze als Teilnahmen befunden und gelangen in die Entität Maßnahme-Teilnahme. Ein fiktives graphisches Beispiel zur Erläuterung dieses Vorgehens findet sich im Anhang (Anmerkung A4).

### 3. Die ISAAK-Maßnahme

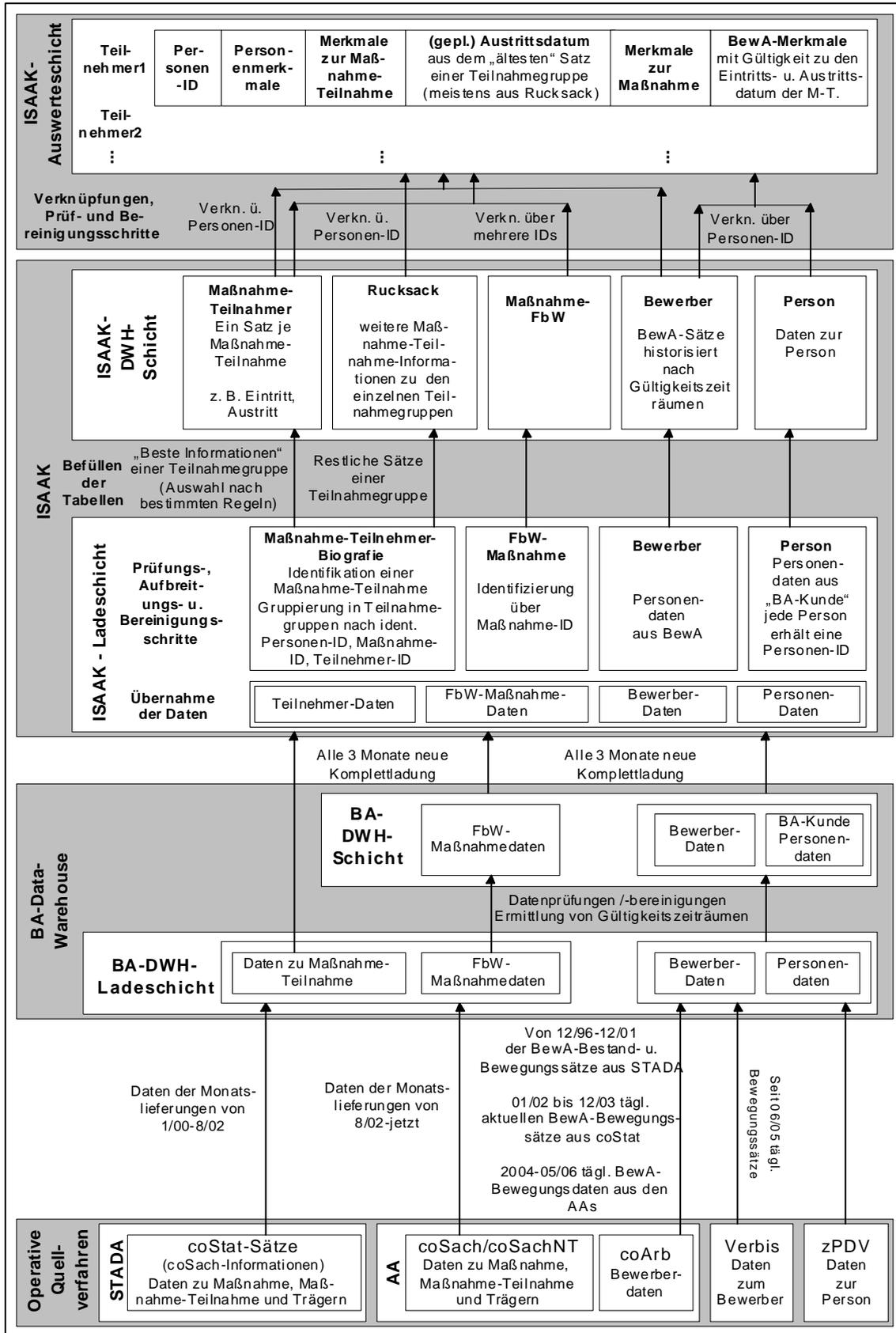
Grundsätzlich werden bei allen Maßnahmen die Maßnahmemerkmale verwendet, die über die coSach(NT)-Sätze zur ISAAK-Maßnahmeteilnahme vorliegen – d.h. die Maßnahmemerkmale werden i.d.R. über die Teilnahmesätze geladen. Ausnahmen bilden die Maßnahmearten FbW und PSA, nur für diese existiert noch eine eigene ISAAK-Maßnahme-Entität.<sup>27</sup>

---

<sup>27</sup> Der Grund hierfür liegt darin, dass für FBW-Maßnahmen – im Gegensatz zu anderen Maßnahmen - mehrere Maßnahmemerkmale vorliegen und dass durch den Aufbau einer eigenen Maßnahmeentität eine effiziente Datenverwaltung ermöglicht wird.

---

Abbildung 4: Aufbau der Datenbank ISAAK und einer Auswerteschicht



## 5.2 Aufbau einer ISAAK-Auswerteschicht

Die ISAAK-Auswerteschicht ist, analog der MTG-BewA-Auswerteschicht, ein Auszug aus ISAAK, bei dem die getrennt gehaltenen Daten der zu untersuchenden Maßnahme-Teilnahmen, der Bewerber und der Person miteinander verknüpft werden. Jeder Teilnahmesatz enthält das Merkmal „geplantes Endedatum“ der Teilnahme. Hierbei handelt es sich um das Austrittsdatum des Maßnahme-Teilnahme-Satzes mit dem niedrigsten Erfassungsdatum (d.h. erste Meldung) einer Teilnahmegruppe. In den Fällen, wo eine Teilnahmegruppe aus nur einem Satz besteht, entspricht das Merkmal „geplantes Endedatum“ dem Attribut „Endedatum“ einer Maßnahme. Bei mehr als einem Satz in einer Teilnahmegruppe, wo in der Regel der aktuellste Satz als Teilnahme angesehen wird, stammt das „geplante Endedatum“ aus dem Satz mit dem niedrigsten Erfassungsdatum (d.h. aus der ersten Meldung) aus der Entität Rucksack der ISAAK-DWH-Auswerteschicht.<sup>28</sup>

Bei der Erstellung einer ISAAK-Auswerteschicht werden zudem die BewA-Merkmale an die Maßnahme-Teilnahme-Merkmale jedes Teilnehmers für zwei verschiedene Zeitpunkte über den ISAAK-Personenidentifikator herangespielt. Dies betrifft den Eintritt (-1 Tag) und den Austritt (+1 Tag). Damit ist gesichert, dass die zu diesen Zeitpunkten gültigen BewA-Merkmale berücksichtigt werden.

## 6. Vergleich einer MTG- und ISAAK-Auswerteschicht

Im Folgenden wird untersucht, ob der unterschiedliche Aufbau der beiden Datenbasen MTG und ISAAK mit Unterschieden in den Maßnahme-Teilnahme-Daten verbunden sind. Hierzu werden Auszüge aus der MTG und aus ISAAK in Form von Auswerteschichten gegenübergestellt. Weiter einschränkend werden nur Teilnahmen an der Existenzgründungsförderung (Coaching) und der beruflichen Weiterbildung (FbW) im Rahmen des ESF-BA-Programms betrachtet, die das gleiche Eintrittsdatum aufweisen. Untersucht werden nur Maßnahme- und Teilnahmemerkmale, deren Quelle bei beiden Datenbasen coSachNT ist bzw. war<sup>29</sup>.

---

<sup>28</sup> Durch diese Regelung kann es beim Vorhandensein von sich überschneidenden Sätzen in einer Teilnahmegruppe dazu kommen, dass das in einem Teilnahmesatz der ISAAK-Auswerteschicht angegebene Endedatum einer Maßnahme vor dem „geplanten Austrittsdatum“ liegt.

<sup>29</sup> Merkmale aus der Bewerberangebotsdatei werden nicht untersucht, da sich hierbei nicht feststellen lässt, ob Veränderungen in deren Ausprägungen auf die Umstellung der Datenbanken oder auf den Wechsel des Fachverfahrens coArb auf VerBIS zurückzuführen sind.

---

### 6.1 Strukturelle Unterschiede und zu erwartende empirische Differenzen

Die wichtigsten Unterschiede zwischen der MTG und ISAAK lassen sich auf drei Aspekte zurückführen: a) Datenziehung und Datenhaltung b) Verknüpfung und Aufbau der Datenstruktur sowie c) Aufbau der Datenfiles.

Ad a) Die MTG basiert im Wesentlichen auf den Strukturvorgaben der STADA-Datenbank und auf den durch coStat gelieferten Datensätzen, die sich jeweils auf unterschiedliche Zeitpunkte beziehen. Bei ISAAK wird dagegen auf die DWH-Struktur der Bundesagentur für Arbeit zurückgegriffen. Durch die technischen Verfahrensänderungen im Zusammenhang mit dem DWH (Ziehungsverfahren, vgl. Punkt 3) sind die Datenbasen zu den Geschäftsstatistiken am aktuellen Rand vor allem vollständiger geworden und qualitativ besser gefüllt (implementierte Prüfroutinen in coSachNT). Weitere Unterschiede begründen sich damit, dass mit ISAAK bisherige Bereinigungs- und Aufbereitungsregeln aufgegeben werden. So wird in ISAAK z.B. grundsätzlich die Regel aufgegeben, dass Teilnahmen nur dann gezählt werden, wenn zu einem Abgangssatz auch Zugangssätze vorliegen. Darüber hinaus hängen die BewA-Meldungen mit coSachNT nicht mehr wie bei coSach an den Teilnahmesätzen, sondern werden separat geführt.<sup>30</sup>

Ad b) Ein gravierender Unterschied zwischen MTG und ISAAK liegt in der verschiedenartigen Architektur der Ausgangsdaten und der Verknüpfung der einzelnen Quellen. Während bei der MTG insgesamt drei zentrale Entitäten unterschieden werden, sind es bei ISAAK nur zwei (Wegfall der Entität Maßnahme, Ausnahme: u.a. FbW, siehe Abschnitt 4.1 / 4.2). Bei der MTG werden die Teilnahmemeldungen künstlich aus den Teilnahmeidentifikatoren und der Relation Maßnahmeidentifikator - Kundenidentifikator gebildet. Dieses Vorgehen führt zum Teil aufgrund nicht eindeutig vergebener Maßnahmenummern zu falschen Teilnahmemeldungen. Bei ISAAK wird auf diese künstliche Identifikation verzichtet.

Ad c) Bei der Erstellung der MTG wurden jeweils neue Meldungen angespielt, während bei ISAAK jeweils komplette Neuladungen der Datenmenge vorgenommen werden. Dies führt dazu, dass mit ISAAK verspätet eingehende Korrekturen berücksichtigt werden und nicht wie bei der MTG ggf. erneute Sätze zur selben Teilnahme existieren. Darüber hinaus werden in ISAAK deutlich weniger Attribute geführt, was zu einer effektiveren Datenhaltung führt.

Diese Unterschiede zwischen ISAAK und MTG lassen erwarten, dass sich systematische Differenzen in der Qualität der Daten mit dem „Systemwechsel“ einstellen. Zunächst ist davon auszugehen, dass sich die Fallzahlen voneinander unterscheiden, wobei zu erwarten ist, dass die Anzahl gültiger Teilnahmen zu einer Maßnahme in ISAAK höher ausfällt als in der MTG.

---

<sup>30</sup> Geblieben ist allerdings das Problem bei verzögerten Eingaben. Durch die Verknüpfungsroutinen bei der Zusammenspielung der BewA-Sätze mit den coSach(NT)-Sätzen kann es weiterhin vorkommen, dass ältere Teilnahme-Sätze mit zeitlich jüngeren BewA-Merkmalen kombiniert werden.

Auch ist zu vermuten, dass sich durch die Abschaffung der Unterscheidung von Sätzen zu unterschiedlichen Zeitpunkten und das neue Ladevorgehen auch Veränderungen in den Datumswerten ergeben können. Ebenso könnten aber auch Unterschiede in den Merkmalsanteilen auftreten. Dies gilt insbesondere dann, wenn man unterstellt, dass Korrektursätze mit einer höheren Wahrscheinlichkeit bei Personengruppen vorliegen, die häufig Leistungen der aktiven Arbeitsmarktpolitik in Anspruch nehmen.<sup>31</sup>

## 6.2 Analysedaten und Datenaufbereitung

Aufgrund der unterschiedlichen Struktur der beiden Datenbasen MTG und ISAAK lässt sich ein einfacher Vergleich zweier Auswerteschichten nicht ohne eine Homogenisierung beider Analysedaten durchführen. Beide Datensätze enthalten Teilnahmemeldungen zu zwei selektierten Maßnahmenarten, deren Teilnahmebeginn zwischen dem 01.01.03 und dem 30.06.03 liegt.

In der MTG existieren jedoch zu jedem Attribut in der Regel zwei Werte: einer zum Zugang (Zugangssatz) und einer zum Abgang (Abgangssatz). In der ISAAK-Auswerteschicht liegt, bis auf wenige Ausnahmen, nur ein Wert je Attribut vor. Um eine adäquate Gegenüberstellung der beiden Auswerteschichten zu gewährleisten, wird bei der Auswahl der Vergleichssätze aus der MTG-Auswerteschicht die Werteausprägung zum Abgang betrachtet.<sup>32</sup> Da bei ISAAK für eine Teilnahme in der Regel der aktuellste Teilnahmesatz zur Datenbildung verwendet wird, müsste dieser in den meisten hier betrachteten Fällen dem Abgangssatz in der MTG entsprechen. Falls diese Variable keine Angabe hat wird das Eintrittsdatum zum Zugang als Auswahlkriterium herangezogen. Außerdem wird wie bei ISAAK innerhalb einer Teilnahmegruppe nur der aktuellste Satz verwendet.<sup>33</sup>

Für die ISAAK-Auswerteschicht wird die Version 2.00 mit aktuellem Datenstand zum März 2006 verwendet. Die MTG-Auswerteschicht basiert dagegen auf der letzten Version zum aktuellen Stand März 2005. Insgesamt umfasst der erstellte Datensatz 31.361 Sätze wovon 15.904 aus ISAAK stammen und 15.457 über die MTG zugespült wurden.

---

<sup>31</sup> Auswertungen zur Qualitätssicherung zwischen ISAAK und MTG untermauern diese Vermutungen zu den Unterschieden zwischen ISAAK und MTG (vgl. Hummel, 2005 und 2006).

<sup>32</sup> Nicht alle Variablen in der MTG-Auswerteschicht sind zweimal vertreten. Es gibt Variablen, deren Merkmalswerte nur aus dem Zugangssatz oder nur aus dem Abgangssatz enthalten sind. Zudem gibt es Teilnahmen, bei denen nur ein Zugangssatz vorhanden ist, d.h. dass alle Variablen, welche sich auf den Abgangssatz beziehen, keine Werte enthalten und nur die entsprechenden Variablen des Zugangssatzes gefüllt sind.

<sup>33</sup> Darüber hinaus werden weitere Anpassungen vorgenommen, die sich auf Angleichung der Variablenamen, des Speichertyps und die Wertebezeichnungen – einschließlich der Missing Values beziehen.

---

### 6.3 Abgleich der Anzahl Kundennummern und Anzahl Sätze/Teilnahmen

Die Eckdaten sind getrennt nach Maßnahmeart und Quelle in Tabelle 1 gegenübergestellt. Es zeigt sich, dass bei beiden Maßnahmearten die Anzahl der Sätze bzw. der Teilnahmen als auch die Anzahl der unterschiedlichen Kundennummern in der ISAAK-Stichprobe höher sind als in der entsprechenden MTG-Stichprobe.

**Tabelle 1: Eckdaten der Stichproben**

	Quelle:	MTG-Auswerteschicht	ISAAK-Auswerteschicht
Coaching	Sätze/Teilnahmen:	9.724	9.922
	Kundennummern:	9.692	9.888
	gemeinsame Kundennummern:	9.669	
	Anzahl Sätze der gem. Kundennr.:	9.701	9.702
ESF-FbW	Sätze/Teilnahmen:	5.733	5.982
	Kundennummern:	5.556	5.623
	gemeinsame Kundennummern:	5.544	
	Anzahl Sätze der gem. Kundennr.:	5.721	5.891

Anmerkung: Ein Vergleich mit den Monitoring-Zahlen der ESF-BA-Begleitforschung liefert andere Zahlenwerte. Zu berücksichtigen ist jedoch, dass hier nur das erste Halbjahr betrachtet wird und dass die Teilnahmen in 2003 nicht gleichmäßig über das Jahr verteilt sind.

Hierfür könnte es zwei bereits in den allgemeinen Bewertungen angesprochene Erklärungsansätze geben. Erstens kommen nur Teilnahmen in die MTG-Auswerteschicht für die coStat einen Zugangssatz liefert, d. h. Teilnahmen für die nur ein Bestands- und/oder Abgangssatz vorhanden war, wurden hier nicht berücksichtigt.<sup>34</sup> Da es bei der Genese von ISAAK keine Differenzierung zwischen Zugangs-, Bestands- und Abgangssatz gibt, werden hier auch die eben beschriebenen Teilnahmen berücksichtigt, die in der MTG herausgefallen sind. Dies würde den oben getroffenen Erwartungen entsprechen. Zweitens könnten die Differenzen in den Fallzahlen auch durch den unterschiedlichen Datenstand der beiden Datenbasen eine Rolle spielen. Die verwendete ISAAK-Datenbank hat einen um ein Jahr aktuelleren Stand als die genutzte MTG-Datenbank. Dies kann zur Folge haben, dass zwischenzeitliche Korrekturen bzw. Nachmeldungen in den Arbeitsagenturen nun in ISAAK berücksichtigt worden sind.

Diese Gründe könnten auch erklären, warum bei Coaching 219 Kundennummern (mit 220 Sätzen) und bei ESF-FbW 79 Kundennummern (mit 91 Sätzen) in ISAAK nicht aber in der MTG-Stichprobe enthalten sind. Weitere 23 fehlende Kundennummern bzw. Sätze in ISAAK, die in der MTG vorhanden sind (bei ESF-FbW betrifft dies 12 Kundennummern bzw. Sätze) könnten auf

<sup>34</sup> Bedenke: Die Informationen kommen aus den Abgangssätzen, Sätze werden in der MTG nur dann berücksichtigt, wenn zu den Abgangssätzen auch Zugangssätze existieren.

nachträgliche Korrekturen am Eintrittsdatum, welches hier das Ziehungskriterium war, zurückzuführen sein, so dass diese Teilnahmen in ISAAK aus dem betrachteten Zeitraum gefallen sind.

Die Tabelle 2 stellt die Verhältnisse der Satzanzahl je Kundennummer zwischen der MTG- und ISAAK-Stichprobe dar. Das Verhältnis 1/2 bedeutet beispielsweise, dass das Ereignis „eine Kundennummer in der MTG einmal und in ISAAK zweimal vorhanden“ bei Coaching 3-mal und bei ESF-FBW 155-mal zu beobachten ist (vgl. Tabelle 2:  $m/i$  ; mit  $m=i$  bedeutet: ebenso viele MTG-Sätze wie ISAAK-Sätze;  $m > i$  bedeutet mehr MTG-Sätze als ISAAK-Sätze et vice versa). Hierbei zeigt sich, dass bei der Maßnahmeart Coaching 99,4 % der Kundennummern aus der MTG die gleiche Anzahl an Sätzen in der MTG als auch in ISAAK haben. Bei den ESF-FbW-Teilnahmen sind dies nur 92,6%. Insgesamt betrachtet ergibt sich damit ein Anteil von 97,4 % zu dem die Kundennummern aus der ISAAK-Stichprobe, die gleiche Anzahl von Teilnahmen in der MTG aufweisen.

**Tabelle 2: Häufigkeiten unterschiedlicher Meldeanzahlen**

Verhältnis der Anzahl Sätze einer Kundennummer in der MTG zu der in ISAAK ( $m/i$ )	Häufigkeiten einer Kundennummer	
	Coaching	ESF-FbW
1/1	9.634	5.214
2/2	30	156
3/3	-	4
0/1	218	68
0/2	2	10
0/3	-	1
1/2	3	155
1/3	-	2
1/4	-	2
2/3	-	6
3/4	-	2
1/0	23	12
2/1	2	3

Bei der Maßnahme ESF-FbW sehen die Zahlen wie folgt aus: 93,7% der in der MTG vorhandenen Kundennummern haben die gleiche Anzahl an Sätzen in der MTG- und ISAAK-Stichprobe, dagegen sind es nur 89,8% der Kundennummern aus der ISAAK-Datei, welche die gleiche Anzahl an Sätzen in der MTG und ISAAK aufweisen<sup>35</sup>.

#### 6.4 Abgleich der Maßnahme- und Teilnahme-Merkmale

Die Untersuchung der Variablen auf Verzerrungen in den Verteilungen der Merkmalsausprägungen erfolgt in einem ersten Schritt für jede Maßnahmeart getrennt mittels eines T-Tests. In einem

<sup>35</sup> Dabei ist zu beachten, dass nur die Variable Kundennummer betrachtet wurde. Das heißt nicht, dass wenn eine Kundennummer einen Satz in MTG und in ISAAK hat, dieser in allen hier betrachteten Variablen übereinstimmen muss.

zweiten Schritt werden dann mittels Kolmogorow-Smirnow-Tests (KS-Test) Untersuchungen auf Gleichheit der Verteilungen durchgeführt (vgl. u.a. Büning/Trenkler, 1994). Die Tabelle 3 gibt hierzu einen ersten Überblick. Aufgeführt sind hier die Anzahl der Variablen, die in die jeweiligen Tests eingebunden wurden, getrennt nach den beiden untersuchten Maßnahmearten Coaching und ESF-FbW.<sup>36</sup>

**Tabelle 3: Überblick der Test-Ergebnisse**

Test	T-Test		Kolmogorow-Smirnow-Test	
	Anzahl an Variablen			
Variablen	getestet	signifikant*	getestet	signifikant *
Coaching	25	6	8	3
ESF-FbW	42	6	12	2

\* signifikant auf 90%-Niveau; dichotome Variablen bleiben unberücksichtigt

Die Ergebnisse im Überblick zeigen, dass sich beim Coaching bei 6 von 25 Variablen über den T-Test statistisch signifikante Unterschiede zwischen der MTG- und der ISAAK-Stichprobe ergeben. Bei ESF-FbW-Förderfällen ist dies ebenfalls bei 6 Variablen der Fall (von 42 getesteten). In die zweiten Tests sind weniger Merkmale einbezogen. Dies ist zurückzuführen auf die unterschiedlichen Skalenniveaus der Variablen. Beim Coaching wurden 8 Variablen bei den Kolmogorow-Smirnow-Verteilungstests berücksichtigt, bei ESF-FbW 12 Variablen. Statistisch signifikante Unterschiede ergeben sich hier beim Coaching bei drei Variablen und bei der ESF-FbW-Förderung bei 2 Variablen.

#### 6.4.1 Ergebnisse bei der Förderung beruflicher Weiterbildung (ESF-FbW)

Im Folgenden werden die Ergebnisse der durchgeführten Tests für die Variablen der Maßnahmeart ESF-FbW genauer dargestellt. Hierbei beschränkt sich die Darstellung auf die Merkmale mit den stärksten Abweichungen in den Merkmalsanteilen bzw. der Verteilungseigenschaften. Die restlichen, ebenfalls statistisch signifikanten Variablen sind inklusive ihrer Ergebnisse im Anhang (Tabelle B1 u. Abbildungen B1 und B2) aufgeführt.

Tabelle 4 beschreibt die Eigenschaften der Attribute Bearbeitungsstatus<sup>37</sup> und der Gesamtdauer<sup>38</sup>. Dargestellt sind hier die Stichprobengröße, der Mittelwert, die Standardabweichung, der

<sup>36</sup> Nicht mit einbezogen bei den verschiedenen Tests wurden die Variablen Eintritts- und Austrittsdatum zur Maßnahmeteilnahme, da diese Variablen dazu verwendet wurden, die hier betrachteten Stichproben zu ziehen. Dies hat zur Folge, dass mögliche vorhandene Streuungen der Merkmalsausprägungen zwischen den Stichproben nicht beobachtbar sind, wenn die entsprechenden Datumsangaben außerhalb des hier betrachteten Zeitfensters liegen und diese Sätze bei der Stichprobenziehung somit unberücksichtigt geblieben sind

<sup>37</sup> Der Bearbeitungsstatus gibt an, ob die Maßnahme bewilligt wurde oder sich noch in Bearbeitung befindet, d.h. der Vorgang noch geprüft wird.

Variationskoeffizient sowie die Schiefe – getrennt ausgewiesen nach Quelle (MTG oder ISAAK). Ebenfalls in Tabelle 4 dargestellt sind die Ergebnisse der vergleichenden Tests bezüglich der Mittelwerte (T-Test) und der Verteilungseigenschaften (KS-Test) der einzelnen Merkmale. Um die Varianz der Merkmale über die beiden Stichproben miteinander vergleichen zu können, wurde der Variationskoeffizient gebildet. Er stellt eine Art normierte Varianz dar, welche den Mittelwert zur Normierung einbezieht.

**Tabelle 4: Eigenschaften zu Bearbeitungsstatus und Gesamtdauer**

	Quelle	N	Mittelwert	Standard Abweichung	Variationskoeffizient	Schiefe	T-Test	KS-Test
<b>Bearbeitungsstatus</b> <sup>1</sup>	MTG	5,733	0.49	0.500	1.014	0.27	x***	-
	ISAAK	5,982	0.82	0.386	0.472	-1.65		
<b>Gesamtdauer (in Monaten)</b> <sup>2</sup>	MTG	5,733	20.97	16.860	0.804	1.11	x***	x (n.s.)
	ISAAK	5,982	22.36	19.656	0.879	1.59		

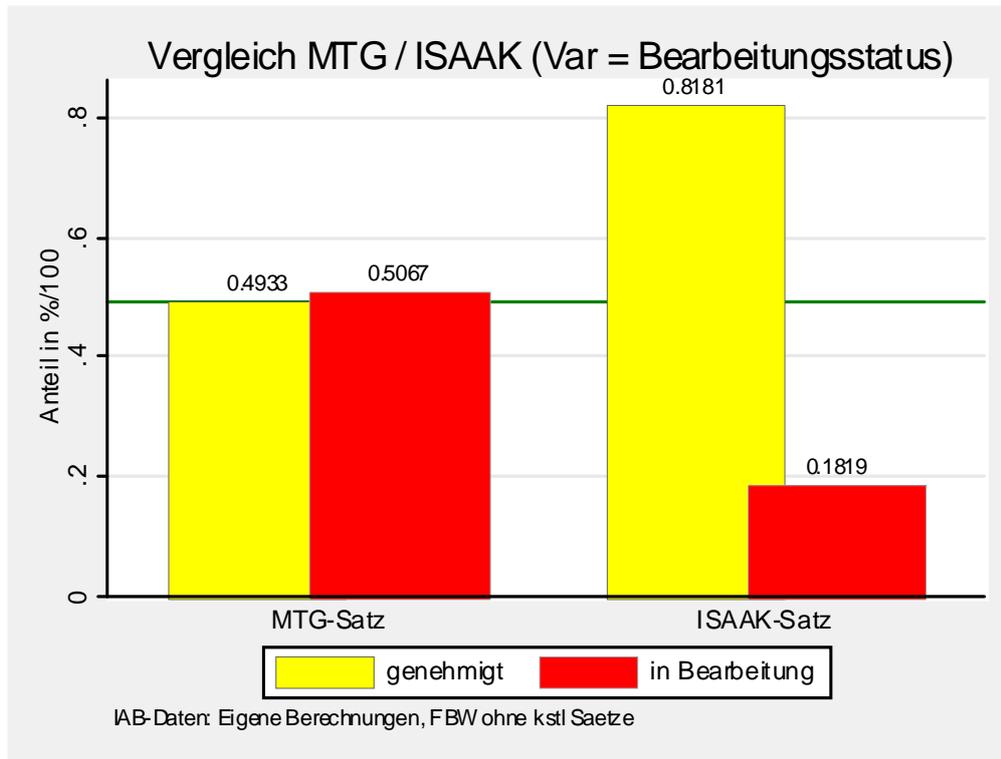
\*\*\*signifikant auf 99%-Niveau \*\* 95%-Niveau \* 90%-Niveau ; n.s. = statistisch nicht signifikant

x = getestet, - = nicht getestet

<sup>1</sup> nominal skaliert <sup>2</sup> metrisch skaliert

Der T-Test auf Unterschiede in den Mittelwerten der Stichproben liefert für die Variable Bearbeitungsstatus ein statistisch hochsignifikantes Ergebnis. Die Mittelwerte weichen mit 0.49 (Anteil genehmigter Teilnahmemeldungen von 49%) bei der MTG-Stichprobe und 0.82 bei der ISAAK-Stichprobe sehr stark voneinander ab. Darüber hinaus ist die Streuung beim Bearbeitungsstatus in der MTG-Stichprobe deutlich größer als in der ISAAK-Stichprobe (vgl. Variationskoeffizient und Standardabweichung). Gleichzeitig macht die hohe Schiefe in der ISAAK-Stichprobe auch darauf aufmerksam, dass der hohe Mittelwert durch eine stark rechtschiefe Verteilung verursacht ist. Insgesamt deuten die Ergebnisse darauf hin, dass in MTG die Anteile der Werte relativ gleich verteilt sind, wogegen sich die Anteile in der ISAAK-Stichprobe vergleichsweise stark konzentrieren.

<sup>38</sup> Die Gesamtdauer der Maßnahme stellt nicht die Maßnahmeteilnahmedauer eines einzelnen Teilnehmers dar, sondern ist ein Maßnahmemerkmals, welches sich auch nur auf die jeweilige Maßnahme bezieht.

**Abbildung 5: Vergleich von Anteilswerten**

Dies lässt sich auch in der graphischen Darstellung des Vergleichs der Merkmalsanteile erkennen. In Abbildung 5 sind die Anteile der Merkmalsausprägungen „genehmigt“ und „in Bearbeitung“ des Bearbeitungsstatus in %/100 getrennt nach Datenquelle dargestellt. Während in der MTG-Stichprobe der Anteil „genehmigt“ markierter Fälle nur leicht geringer ist als der in „in Bearbeitung“, kehrt sich dieses Verhältnis in der ISAAK-Stichprobe um. In den Eigenschaften der Verteilung spiegelt sich dies in der Schiefe, mit Werten von 0.27 auf MTG- und -1.56 auf ISAAK-Seite, wieder.

Bei näherer Betrachtung der Daten wurde festgestellt, dass dieser Unterschied durch den aktuelleren Datenstand von ISAAK zustande kommt. Da der Bearbeitungsstatus eine zeitvariante Variable ist, ergeben sich trotz der schon relativ lang angesetzten Nachlaufzeit (Zeit zwischen betrachteten Zeitfenster und dem Ladedatum der Datenbanken) von 2 Jahren in beiden Analysendatensätzen derartige Verschiebungen.

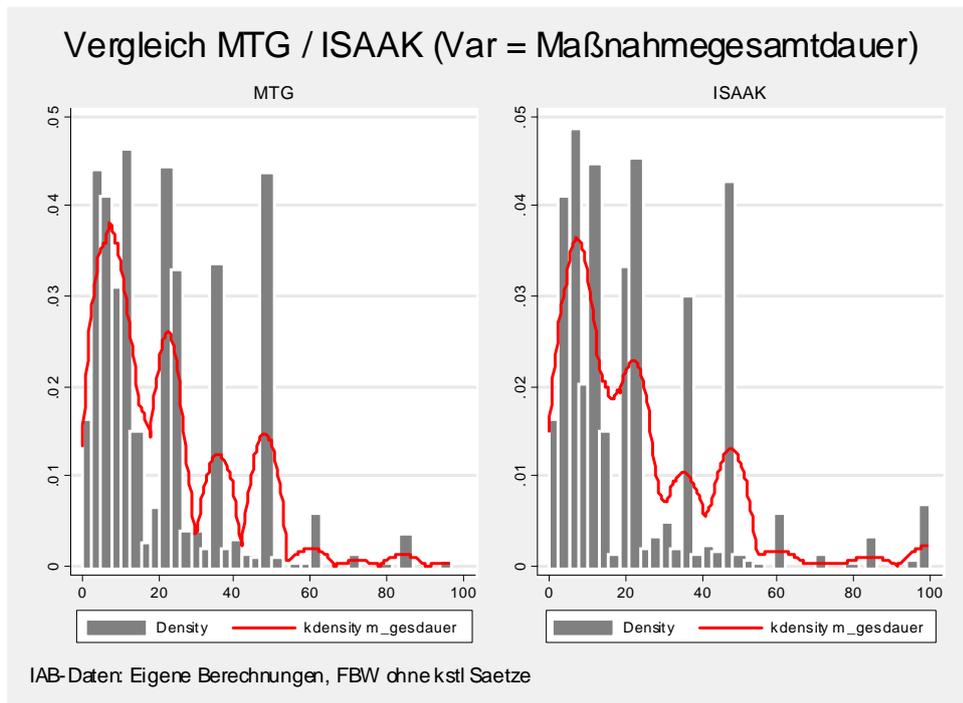
Die Variable „Gesamtdauer“ (Maßnahmedauer)<sup>39</sup> weist ebenfalls einen statistisch hochsignifikanten Unterschied in den Mittelwerten auf. Dies bedeutet, dass bei einer mittleren Dauer von fast 21 Monaten in der MTG-Stichprobe und 22,36 Monaten in der ISAAK-Stichprobe, die Abweichung 1,39 Monate beträgt. Hier steigt allerdings die Standardabweichung mit dem Mittelwert, so dass ein Vergleich der Variationskoeffizienten, mit 0.804 und 0.879, letztlich keinen so starken Unterschied bedeutet. Der Kolmogorow-Smirnow-Test ist durchführbar. Gestestet wird die Übereinstimmung zweier Wahrscheinlichkeitsverteilungen. Für die Dauer ist er auf dem üblichen Niveau von mindestens 90% nicht statistisch signifikant.

Die Abbildung 6 zeigt hierzu die Dichteverteilung als auch die geschätzte Kern-Dichteverteilung der Maßnahmegesamtdauer für die Daten aus der MTG-Stichprobe (linke Teilabbildung) und aus der ISAAK-Stichprobe (rechte Teilabbildung). Auf der Abszisse sind die Monate der Maßnahmegesamtdauer abgetragen. Der Verlauf der Kerndichteverteilung zwischen MTG und ISAAK weicht nicht stark voneinander ab. Lediglich gegen Ende des Wertebereichs, bei Dauer gegen 100 Monate, deutet sich bei ISAAK ein Anstieg an. Diese ist in den Daten der MTG nicht zu erkennen. Prinzipiell ist zu sehen, dass bei den Daten aus der MTG die „Schluchten“ (lokale Minima) zwischen den Peaks (lokale Maxima) stark ausgeprägt sind, wogegen bei ISAAK der Graph weicher verläuft. Grob betrachtet verweisen diese Unterschiede darauf, dass in der MTG mehr kürzere Teilnahmedauern vorhanden sind. Dies zeigt sich vor allem bei der stärkeren Konzentration von kurzen Dauern in der MTG-Stichprobe aber auch darin, dass lange Dauern hier nur sehr selten auftreten. Außerdem wird deutlich, dass die Maßnahmedauern bei der MTG stärker „klumpen“ als in den Daten der MTG.

---

<sup>39</sup> Zu beachten ist, dass die Variable Maßnahmedauer sich nicht auf die Dauer der Teilnahme an einer Maßnahme bezieht, sondern auf die Dauer der Maßnahme. Im Regelfall sind diese beiden Angaben identisch und liegen im Maximum bei 24 Monaten. Abweichungen beider Dauer lassen sich unter anderem durch Fehleingaben erklären. Andererseits gibt es aber auch Maßnahmen mit laufendem Einstieg, so dass die Maßnahmedauern z.T. deutlich längere Zeiten aufweisen als die Teilnahmen. Bei Dauern länger als 100 Tage kann davon ausgegangen werden, dass es sich um Fehleingaben handelt.

---

**Abbildung 6: Vergleich von Verteilungen**

#### 6.4.2 Ergebnisse bei der Maßnahmeart Coaching

Die Ergebnisse des Vergleichs der Lageparameter und der Tests der Variablen der Maßnahmeart Coaching sind für ausgewählte Merkmale (Basisförderung<sup>40</sup>, Maßnahmedauer in Tagen) in Tabelle 5 dargestellt. Weitere Ergebnisse sind im Anhang in Tabelle B2 und Abbildungen B4 – B6 aufgeführt. Wiederum sind die Angaben zur Größe der Stichprobe, Mittelwert, Standardabweichung, Variationskoeffizient und Schiefe pro Merkmal ebenso wie bei der Maßnahmeart FbW je nach Quelle in MTG- und ISAAK unterteilt.

Bei der Basisförderung zum Coaching wird zwischen dem Erhalt eines Überbrückungsgeldes und eines Existenzgründungszuschusses differenziert. Wie man Tabelle 5 entnehmen kann, ist die Stichprobe aus ISAAK 1,5 Mal so groß wie die der MTG. Neben diesem Unterschied in der Stichprobengröße liegen die Werte der betrachteten Lage- und Verteilungsmaße jedoch relativ dicht bei einander. Der Mittelwert ist mit einer Differenz von 0.05 nur minimal größer und auch die Standardabweichung ist nur minimal kleiner (-0.032). Statistisch gesehen ist dieser Unterschied in der Verteilung jedoch signifikant von Null verschieden.

<sup>40</sup> Die Variable Basisförderung gibt an, ob der Teilnehmer Überbrückungsgeldempfänger war oder einen Existenzgründungszuschuss bekam.

**Tabelle 5: Eigenschaften zu Basisförderung und Maßnahmedauer**

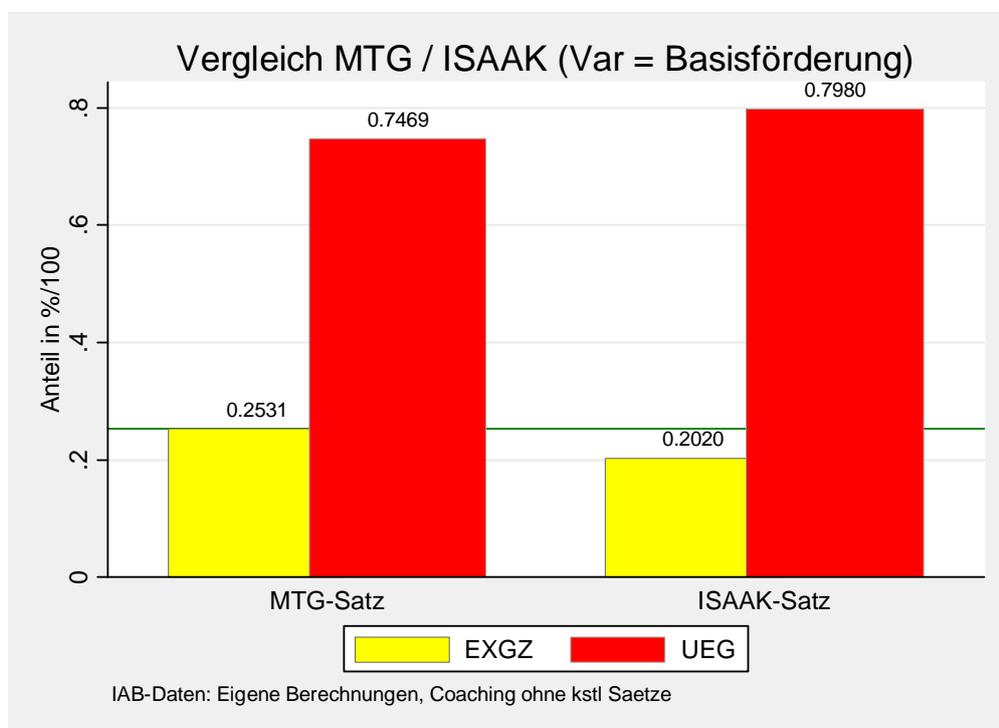
	Quelle	N	Mittelwert	Standard Abweichung	Variationskoeffizient	Schiefe	T-Test	KS-Test
<b>Basis-Förderung<sup>1</sup></b>	MTG	2,114	0.75	0.434	0.582	-1.14	x***	-
	ISAAK	3,169	0.80	0.402	0.503	-1.48		
<b>Maßnahmedauer (in Tagen)<sup>2</sup></b>	MTG	9,724	909.52	408.628	0.449	1.50	x***	x***
	ISAAK	9,922	942.72	393.787	0.418	1.61		

\*\*\*signifikant auf 99%-Niveau \*\* 95%-Niveau \* 90%-Niveau

x = getestet, - = nicht getestet

<sup>1</sup>nominal skaliert, gibt an, ob ÜG oder EXGZ als Basisförderung vorliegt <sup>2</sup>metrisch skaliert

Die graphische Gegenüberstellung veranschaulicht diesen Unterschied nochmals (Abbildung 7). Bei beiden Stichproben steht der Großteil des Coachings (75% bei der MTG-Stichprobe und 80% in der ISAAK-Stichprobe) im Zusammenhang mit einer Förderung zum Überbrückungsgeld. Eine genauere Betrachtung zeigt, dass sich der Unterschied aus den über 1000 Sätzen der ISAAK-Stichprobe ergibt, welche in der MTG nicht enthalten sind. Diese sind fast vollständig Personen mit Überbrückungsgeld.

**Abbildung 7 Vergleich von Anteilswerten**

Die Werte der Lage und Verteilungsmaße zu der Variablen Maßnahmedauer verhalten sich ähnlich wie die der Basisförderung: In der ISAAK-Stichprobe zeigt sich ein größerer Mittelwert sowie eine geringere Standardabweichung. Beide Tests sind statistisch hochsignifikant (T-Test; KS-Test). Trotz der Existenz von statistisch signifikanten Unterschieden lassen sich diese in der graphischen Darstellung nicht erkennen. Da die Abweichungen materiell gering erscheinen, wird eine Erklärung der Unterschiede an dieser Stelle nicht vorgenommen.

#### 6.4.3 Erklärungsansatz

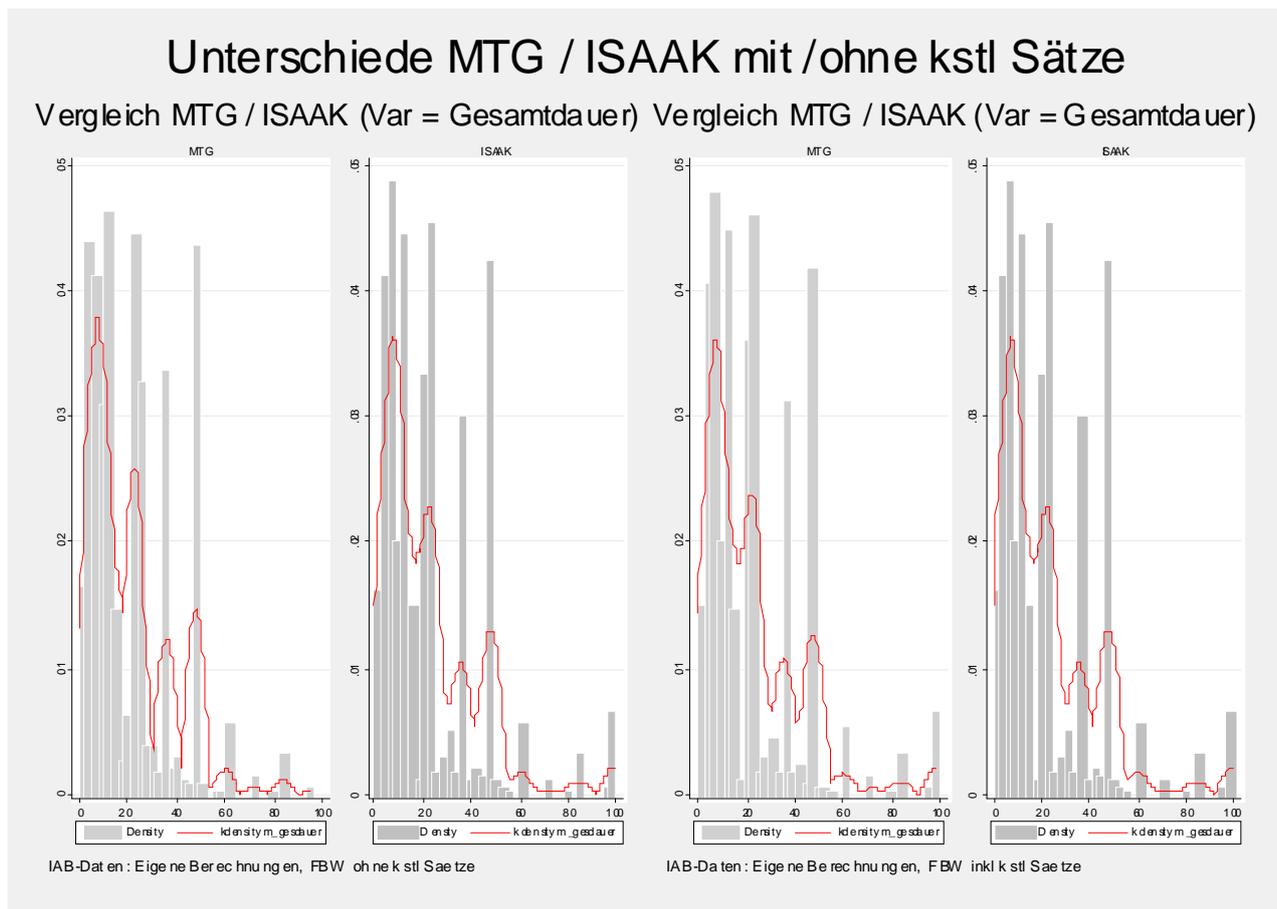
Sowohl die Komplexität der Veränderungen in der Datengenese als auch die Vielzahl an technischen Unterschieden in der Datenaufbereitung, in den Prüfroutinen und in der Verknüpfung der Daten macht es schwierig, detaillierte Erklärungen für einzelne Verteilungsunterschiede in den Stichproben zu geben. Stattdessen werden im Nachfolgenden allgemeine Erklärungsmuster vorgestellt, welche die hier gefundenen Unterschiede beschreiben und erläutern helfen.

Wie schon weiter oben beschrieben, wurden in der MTG die Sätze zu Teilnahmen mit Markierungen für Zugang, Bestand, Abgang und Verbleib versehen. Gab es zu einer Teilnahme keinen Zugangssatz, dann wurde diese Teilnahme nicht in die Auswerteschicht übernommen. In ISAAK existiert diese Unterscheidung der Sätze nicht mehr, womit auch Teilnahmen in die Auswerteschicht gelangen, die früher herausgefiltert wurden. Dies ist eine mögliche Erklärung für die unterschiedlichen Fallzahlen und somit auch für die Verschiebungen in den Ausprägungen der einzelnen Merkmale. Das Projekt zur Begleitforschung des ESF-BA-Programms versuchte mit Hilfe von ITM diese Sätze in der MTG zu „retten“ und bildete einen künstlichen Zugangssatz, falls dieser nicht vorhanden war. Durch das Einbeziehen dieser so genannten „künstlichen Sätze“ in der MTG-Stichprobe verändert sich das Verhältnis der Fallzahlen zwischen den Stichproben. Die Satzanzahl der MTG bei FbW erhöht sich um 365 Fälle auf 6.098, bei 5.982 Sätzen in ISAAK. Etwas geringer ist die Zahl der künstlichen Sätze bei der Maßnahmeart Coaching, hier betrifft dies 119 Sätze, welche nun in der MTG-Stichprobe berücksichtigt werden. Somit ergibt sich ein MTG/ISAAK-Fallzahlen-Verhältnis von 9.843/9.922. Bei beiden ist zu erkennen, dass sich das Verhältnis fast ausgleicht.

Zur Illustration sind in der folgenden Graphik (Abbildung 8) daher die Dichteverteilung und die geschätzte Kern-Dichte-Verteilung der Gesamtdauer in Monaten von ESF-FbW-Maßnahmen exemplarisch abgebildet. Auf der linken Seite sind die Ergebnisse für die Daten aus MTG und ISAAK dargestellt ohne die künstlichen Sätze einzubeziehen, auf der rechten Seite wurden diese dagegen in der MTG berücksichtigt. Wie schon an anderer Stelle beschrieben, sind die Schluchten (lokale Minima) des Graphs der MTG-Daten ohne die künstlichen Sätze tiefer als die des

ISAAK-Graphs (stärkere Klumpung der Dauern). Betrachtet man nun den Graph der MTG-Daten unter Einbeziehung der künstlichen Sätze, wird deutlich, dass die Unterschiede kaum noch vorhanden sind.

**Abbildung 8 Vergleich von Verteilungen – Erklärungsansatz**



## 7. Zusammenfassung und Fazit

Der vorliegende Beitrag dokumentiert den Prozess der Datengenerierung zu Maßnahmeteilnahmen - insbesondere zum ESF-BA-Programm - und stellt empirisch heraus, ob sich durch die Veränderung der Datenquelle der Teilnahmen zu den Maßnahmen der aktiven Arbeitsmarktförderung potentielle Datenbrüche ergeben. Derartige Datenbrüche hätten erhebliche Auswirkungen auf die Validität der Daten für Analysezwecke z.B. den Evaluationsergebnissen zu Maßnahmen und Leistungen der aktiven Arbeitsmarktpolitik.

Die Arbeit macht deutlich, dass die Genese der Teilnahmemeldungen und der Daten zum Personenprofil komplex sind. Darüber hinaus zeigt sich, dass die Unterschiede in der Aufbereitung, den Prüfroutinen und der Datenverknüpfung zwischen beiden Datenbasen erheblich sind. Während diese in der MTG jeweils zu mehreren Zeitpunkten vorliegen und sukzessiv an den neuen

Datenstand angespielt werden, werden die Sätze in ISAAK regelmäßig neu geladen. Dadurch können bei ISAAK auch verspätete Meldungen berücksichtigt werden, die in der MTG als nicht gültig eingestuft wurden. Dies hat zur Folge, dass sich einerseits Unterschiede zwischen den beiden Verfahren bei zeitvarianten Variablen ergeben, sich andererseits aber auch durch die höhere Anzahl an berücksichtigten Sätzen Unterschiede in den Verteilungen bei einzelnen „statischen“ Variablen ergeben können. Festzuhalten ist aber, dass die vorliegenden Ergebnisse keine Hinweise auf materielle Brüche im Informationsgehalt zwischen den beiden Datenquellen begründen. Die Regeln in der MTG sorgen dafür, dass Teilnahmen vor 2000 ohne Zugangssatz unterschätzt werden.

Unklar ist bisher jedoch, ob bestimmte Personengruppen stärker durch die Systemänderungen betroffen sind oder nicht. Beispielsweise wäre zu vermuten, dass es bei Personen mit häufigem Kontakt zur Bundesagentur für Arbeit – z.B. durch häufige Arbeitslosen- bzw. Arbeitssuchemeldungen oder mit häufigeren Teilnahmen - wahrscheinlicher zu Korrekturen kommt. Derartige Untersuchungen stehen noch aus.

## Literatur

Blaschke, D.; Plath, H.-D.; Nagel, E. (1992): Konzepte und Probleme der Evaluation aktiver Arbeitsmarktpolitik am Beispiel der Fortbildung und Umschulung. Mitteilungen aus der Arbeitsmarkt- und Berufsforschung, Jg. 25, H. 3, S. 381-405.

BMAS (2006): Die Wirksamkeit moderner Dienstleistungen am Arbeitsmarkt – Bericht 2006 des BMAS zur Wirkung der Umsetzung der Vorschläge der Kommission Moderne Dienstleistungen am Arbeitsmarkt (ohne Grundsicherung für Arbeitssuchende), Bundesministerium für Arbeit und Soziales, Berlin.

Büning, H.; Trenkler, G. (1994): Nichtparametrische statistische Methoden, 2. Auflage. Verlag Walther de Gruyter, Berlin / New York.

Caliendo, M.; Hujer, R.; Thomsen, S.L. (2003): Evaluation der Netto-Effekte von ABM in Deutschland. Ein Matchingansatz mit Berücksichtigung von regionalen und individuellen Unterschieden. IAB-Werkstattbericht, Nr. 2 vom 10.03.2003, Institut für Arbeitsmarkt- und Berufsforschung, Nürnberg

Deeke, Axel (2004): Die Umsetzung des ESF-BA-Programms in der Zeit von Anfang 2000 bis Ende 2002. In: Institut für Arbeitsmarkt- und Berufsforschung, Nürnberg (Hrsg.), Halbzeitbewertung des "ESF-BA-Programm 2000-2006", (Beiträge zur Arbeitsmarkt- und Berufsforschung, 283), Nürnberg.

Deeke, A.; Wiedemann, E. (2002): Evaluierung aktiver Arbeitsmarktpolitik und Datengrundlagen. Bericht von einem Workshop in der Bundesanstalt für Arbeit am 9. November 2001. IAB-Werkstattbericht Nr. 2 vom 20.03.2002, Nürnberg.

Deeke, A.; Kruppe, T.; Müller, P.; Schuler, W. (2001): Begleitforschung zum "ESF-BA-Programm 2000-2006" - Erster Zwischenbericht. IAB-Projektbericht, 10-534, Institut für Arbeitsmarkt- und Berufsforschung, Nürnberg

Kruppe, T.; Oertel, M. (2003): Von Verwaltungsdaten zu Forschungsdaten: Die Individualdaten für die Evaluation des ESF-BA-Programms 2000-2006, IAB-Werkstattbericht, Nr. 8/2002, Nürnberg

Schwarzfärber, H. P. (2002): Das Data-Warehouse der BA. In: Deeke, A./Wiedemann E.: Evaluierung aktiver Arbeitsmarktpolitik und Datengrundlagen – Bericht von einem Workshop in der Bundesanstalt für Arbeit am 9. November 2001, IAB-Werkstattbericht, Nr. 2/2002, Anhang 1, Nürnberg

## Quellen

Bundesagentur für Arbeit (2003): Geschäftsanweisung – Flächeneinführung der IT-Anwendung zPDV (zentrale Personendatenverwaltung). BA-Rundbrief 86/2003 vom 14. Juli 2003, Bundesagentur für Arbeit, Nürnberg

Burchard, K.; Ferrari, V.; Hummel, E.; Zitzelsberger, A. (2005): Benutzer-Handbuch-ISAAC (Instrumente aktiver Arbeitsmarktpolitik) Version 0

Dade, C.; Hummel E.; Kaimer, S.; Kellner, E.; Oertel, M.; Passenberger, J.; Rauscher, A.; Schmidt, K.; Ziegerer, M.; Beck, W.; Stamm, M.; Majer, W. (2004): Integrierte IAB-Individualdatenbasis, Version 1.00, Kurzbeschreibung.

Fischer (2002): Fachkonzept Übernahme der BEWA-Daten aus dem Filesystem in die Ladeschicht des Data Warehouse

Hummel, E. (2005): QS-Prüfungen – Vergleich der Auswertefiles aus ISAAC und MTG, Version 1.03 vom 04.08.2005, Institut für Arbeitsmarkt- und Berufsforschung, Nürnberg

Hummel, E. (2006): QS-ISAAC ohne MTG – QS-Prüfungen ISAAC zu Verfahren, die nicht in MTG enthalten sind. Version 1.00 vom 19.10.2006, Institut für Arbeitsmarkt- und Berufsforschung, Nürnberg

Instrumente aktiver Arbeitsmarktpolitik (ISAAC) Anwenderschicht ESF-Maßnahme-Teilnahme Version V2.01 Benutzerhandbuch (2006)

Projekt pallas-iv (2005): ER-Datenschicht ISAAC komplett 041222. Internes Dokument vom 20.01.2005, Institut für Arbeitsmarkt- und Berufsforschung, Nürnberg

Seysen, C. (2006): Erweiterte QS – Ergebnisse des Versionsvergleichs IEB V3.01/V4.02

Seysen, C. (2006): Kontinuität zwischen IEB V3 und V4 in ‚ITM for Users‘ Ausgabe 2 April 2006

Zitzelsberger, A. (2002): Fachkonzept Übernahme der BEWA-Daten aus der Ladeschicht in das Data Warehouse

Zitzelsberger, A. (2005): SA ISAAC 050120.doc. Dokumentation des Modells ‚Pallas-iv versorgen‘ aus dem Repository ‚IAB‘

## Anhang A

Anmerkung A1:

Die Regeln im Einzelnen:

- Ein Datensatz galt als Zugangssatz, wenn das Datum des Eintritts in die Maßnahme zwischen den Stichtagen der letzten und der aktuellen Ziehung lag.
- Bestandsätze waren dadurch definiert, dass das Eintrittsdatum einer Maßnahmeteilnahme vor dem aktuellen Ziehungsstichtag lag und sich das Austrittsdatum noch in der Zukunft befand.
- Bei Abgangssätzen lag das Austrittsdatum der Teilnahme zwischen dem letzten und dem vorletzten Stichtag. Durch diese Regel wurden Abgangssätze mit einem Monat Verzögerung erzeugt, wodurch gering verspätete Eingaben berücksichtigt werden konnten.
- Zur Kontrolle des Verbleibs der Kunden nach Austritt wurden zusätzlich noch Verbleibssätze generiert. Diese basierten auf einem Abgleich mit dem Arbeitslosenstatus und wurden für die Stichtage, einen Monat nach Austritt und 6 Monate nach Austritt erzeugt.

Sowohl Eintritts- als auch Austrittsdatum wurden (und werden) als Planungsdaten der Teilnahme an einer ESF-Maßnahme nach einer Bewilligung „dem Grunde nach“ vor Maßnahmebeginn durch den jeweiligen Sachbearbeiter festgelegt. Unabhängig davon war die elektronische Erfassung in coSach/coSachNT, die zum Teil erst einige Zeit nach der (geplanten) Teilnahme erfolgte. Zum Zeitpunkt der Festlegung (Planung) war die tatsächliche Teilnahme an der Maßnahme insgesamt noch nicht gesichert. Darüber hinaus konnte sich sowohl der Eintritt (z. B. durch Krankheit) als auch der Austritt (z. B. wegen ebenfalls verspätetem Eintritt, vorzeitigem Abbruch oder Verlängerung der Teilnahme) verschieben. Möglicherweise sind nicht alle solche Veränderungen zeitgerecht in die Daten eingeflossen, um bei der Ziehung der Daten für die Geschäftsstatistik Berücksichtigung zu finden. Durch dieses automatisierte Ziehungsverfahren ist nicht überprüfbar, ob die geplante Teilnahme tatsächlich jeweils stattgefunden hat. Gibt es jedoch Abweichungen in den Daten zwischen Zugangs-, Bestands- und Abgangs- bzw. Verbleibssatz, so wurden entsprechende Korrekturen zeitgerecht eingearbeitet. Geht man davon aus, dass der Anteil von nicht erfassten Korrekturen an den für die Statistik relevanten Daten, darunter auch Eintritts- und Austrittsdatum, nicht allzu hoch sein sollte, konnte man die Daten aus dem Zugangsdatensatz generell als „geplant“, die Daten aus dem Abgangssatz als „realisiert“ betrachten.

---

#### Anmerkung A2:

Im Jahre 1998 wurde mit dem Aufbau des Data Warehouse (DWH) der Bundesagentur für Arbeit begonnen. Mit dem BA-DWH sollten u. a. die bestehenden Verfahren coStat/STADA ersetzt werden. Beim BA-DWH handelt es sich um ein Datenbanksystem, in dem die Daten der operativen IT-Verfahren (wie coSach, coArb oder der zPDV) und externen Stellen (z. B. Daten von Sozialversicherungsträgern) besser geprüft, aufeinander abgestimmt und verknüpft werden können. Im Gegensatz zu den in STADA isoliert betrachteten Meldungen zum Zugang, Bestand, Abgang und Verbleib einer Maßnahme werden im DWH die Daten historisiert abgelegt. Somit können die gewünschten Sätze bzw. Konten einfacher erstellt werden, mit denen die individuellen Verknüpfungen von Informationen in zeitlicher Abfolge möglich ist. Das BA-DWH lässt sich vereinfacht in 3 Schichten (Ladeschicht, DWH-Schicht und Auswerteschichten) unterteilen (Details hierzu in: Schwarzfärber, 2002). Die operativen bzw. externen Daten werden in die Ladeschicht geladen und ersten Prüfungen unterzogen. Die Daten werden dann nach weiteren Prüfungen in die so genannte BA-DWH-Schicht entsprechend ihrer Historie integriert. Auswerteschichten (DataMarts) setzen auf spezifische Aufbereitungen der Daten dieser BA-DWH-Schicht als gemeinsame Geschäftsdatenbasis der BA seit 2005 auf und stehen für statistische Auswertungen zur Verfügung.

#### Anmerkung A3: Aufbau der DWH-BewA-Schicht

Die Daten der DWH-BewA-Schicht stammen zu verschiedenen Zeitpunkten aus unterschiedlichen Quellen. Potentielle Datenquellen für die Bewerber- und Personendaten waren das IT-Fachverfahren coArb und die zentrale Personendatenbank CORA bzw. die zentrale Kundendatei (zPDV) der BA. Daten des Zeitraums 12/96 bis 12/01 stammten aus dem Datenbestand an BewA-Bestands- und Bewegungssätzen der zentralen Datenbank STADA. Von 01/02 bis 12/03 lieferten die einzelnen Agenturen täglich mittels coStat die aktuellsten BewA-Bewegungssätze. Seit Anfang 2004 übermitteln nun die Agenturen diese direkt an das DWH. Personendaten stammen nun aus der täglichen Lieferung der zPDV. Die Daten werden in die DWH-Ladeschicht gespeichert und einmal im Monat mit gleichzeitiger Prüfung und einigen Bereinigungen in die DWH-Schicht eingelesen.

Die BewA-Bewegungssätze, welche als eindeutigen Identifikator einer Person die BewA-Eingabenummer und/oder die Kundennummer enthalten, werden nach dem Eingabedatum sortiert und in die bereits bestehenden BewA-Sätze in der DWH-Schicht eingearbeitet. Die historisierte Haltung der BewA-Sätze einer Person erfolgt über das Eingabedatum, wodurch sich ein Gültigkeitszeitraum definieren lässt. Parallel hierzu wird mittels der Personendaten aus der

---

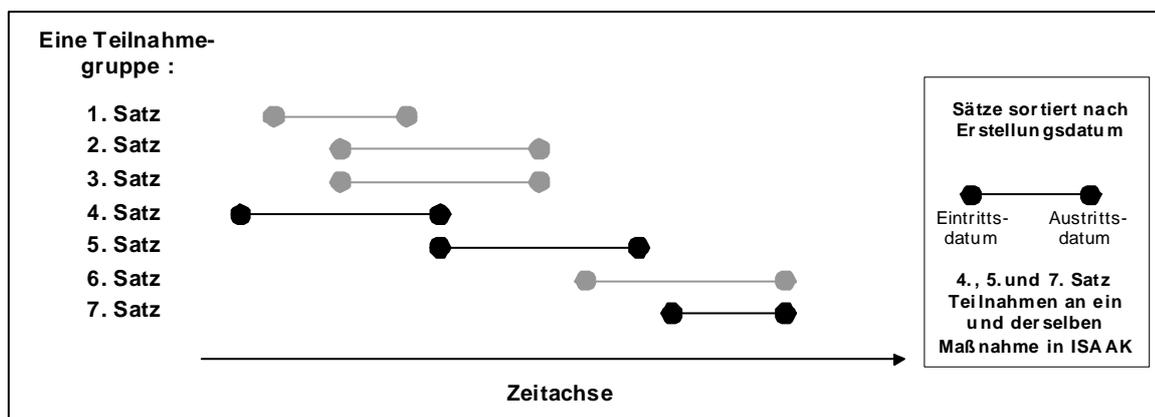
Ladeschicht die Datenbank der BA-Kunden aufgebaut bzw. mit den zeitinvarianten Personendaten der Kunden aktualisiert.

In der DWH-Schicht existieren in den Tabellen zwei eindeutige Identifikatoren für eine Person, die BA-Kundennummer und die BewA-Eingabenummer. Mit Hilfe der BA-Kundennummer wurde in der BewA nach einer zugehörigen BewA-Eingabenummer gesucht. Zu einer Kundennummer kann es daher mehrere BewA-Eingabenummern geben und umgekehrt. In solchen Fällen mit multiplen Personenidentifikatoren wird bei der Verknüpfung zufällig ein Satz selektiert. Die restlichen Sätze bleiben in solchen Fällen unberücksichtigt.

Anmerkung A4: graphisches Beispiel zur Auswahl des „besten“ Satzes

Es handelt sich in Abbildung A1 um eine Teilnahmegruppe die aus 7 Sätzen besteht. Die einzelnen Sätze sind durch Linien dargestellt. Der linke Knoten an einer Linie repräsentiert das Eintrittsdatum und der rechte das Austrittsdatum der Teilnahme wie es in dem einzelnen Satz angegeben ist. Die Zeitachse gibt die Position der Teilnahmezeiträume der einzelnen Sätze zueinander an. Nach den eben beschriebenen Regeln werden der 4., 5. und 7. Satz als Teilnahmen an ein und derselben Maßnahme in die ISAAK-DWH-Schicht übernommen. Die restlichen Sätze werden in dem Rucksack gespeichert.

**Abbildung A1: Gültigkeitszeiträume der Sätze einer Teilnahmegruppe, die nach dem Erstellungsdatum sortiert wurden**



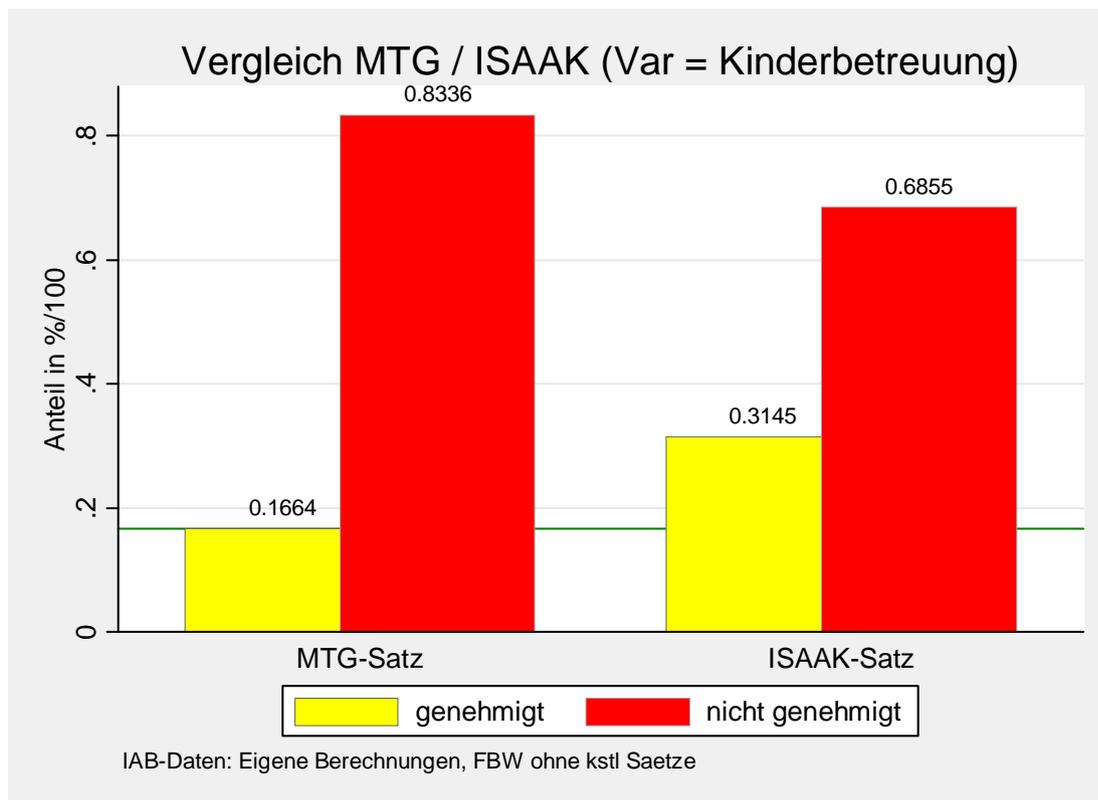
Anhang B

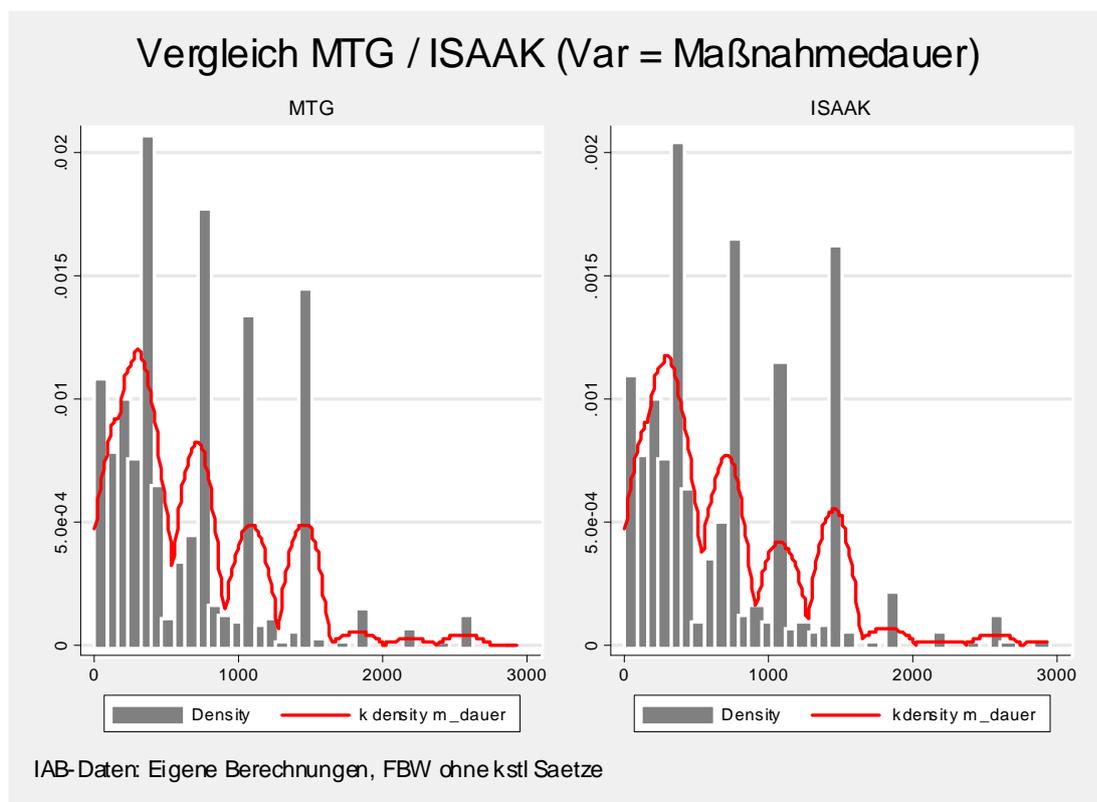
Tabelle B1: Testwerte ESF-FbW

	Quelle	N	Mittel-Wert	Standard Abweichung	Variationskoeffizient	Schiefe	T-Test	KS-Test
Kinderbetreuung	MTG	5,126	0.83	0.372	0.447	-1.79	x***	-
	ISAAK	5,341	0.69	0.464	0.677	-0.80		
Maßnahmedauer (in Tagen)	MTG	5,404	661.34	502.914	0.760	1.02	x*	x**
	ISAAK	5,533	678.79	523.388	0.771	1.00		

\*\*\*signifikant auf 99%-Niveau \*\* 95%-Niveau \* 90%-Niveau  
 x = getestet, - = nicht getestet

Abbildung B1: Vergleich MTG / ISAAK; Anteil mit und ohne Kinderbetreuung (FBW)



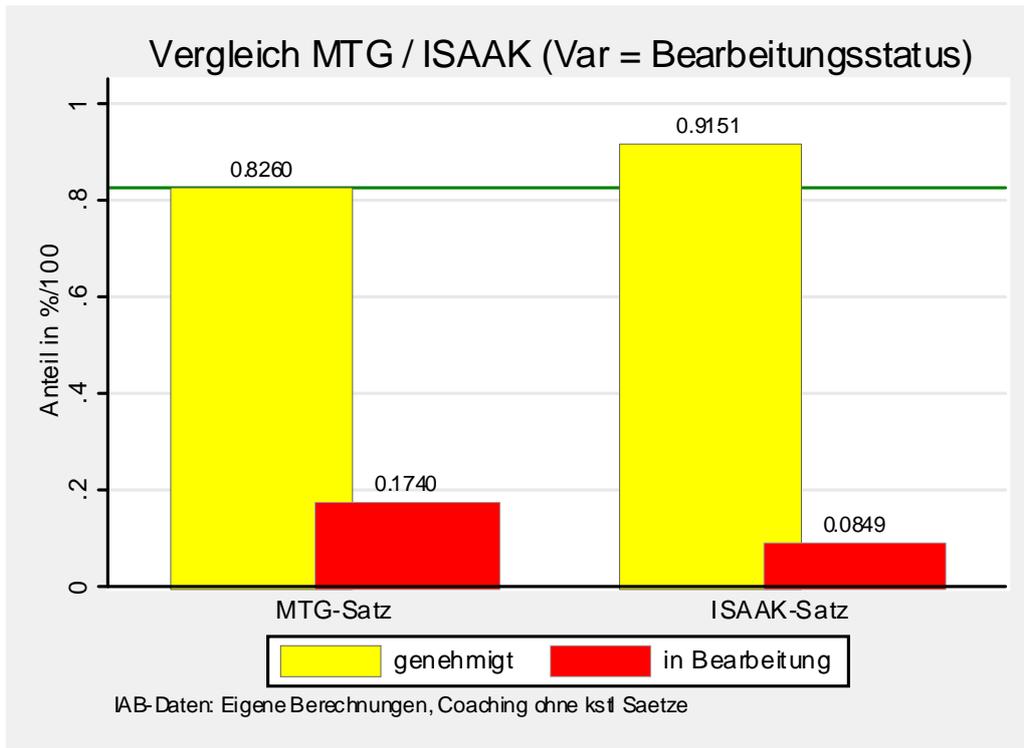
**Abbildung B2: Vergleich MTG / ISAAK; Dichteverteilung Maßnahmedauer (FBW)****Tabelle B2: Testwerte Coaching**

	Quelle	N	Mittelwert	Standard Abweichung	Variationskoeffizient	Schiefe	T-Test	KS-Test
<b>Bearbeitungsstatus</b>	MTG	9,724	0.83	0.379	0.459	-1.72	X***	-
	ISAAK	9,922	0.92	0.279	0.305	-2.98		
<b>Kinderbetreuung</b>	MTG	6,882	0.94	0.229	0.242	-3.88	X***	-
	ISAAK	6,968	0.96	0.188	0.195	-4.93		
<b>Dauer Förderung (in Monaten)</b>	MTG	9,724	28.42	14.376	0.506	1.23	X**	X***
	ISAAK	9,922	28.90	14.509	0.502	1.03		

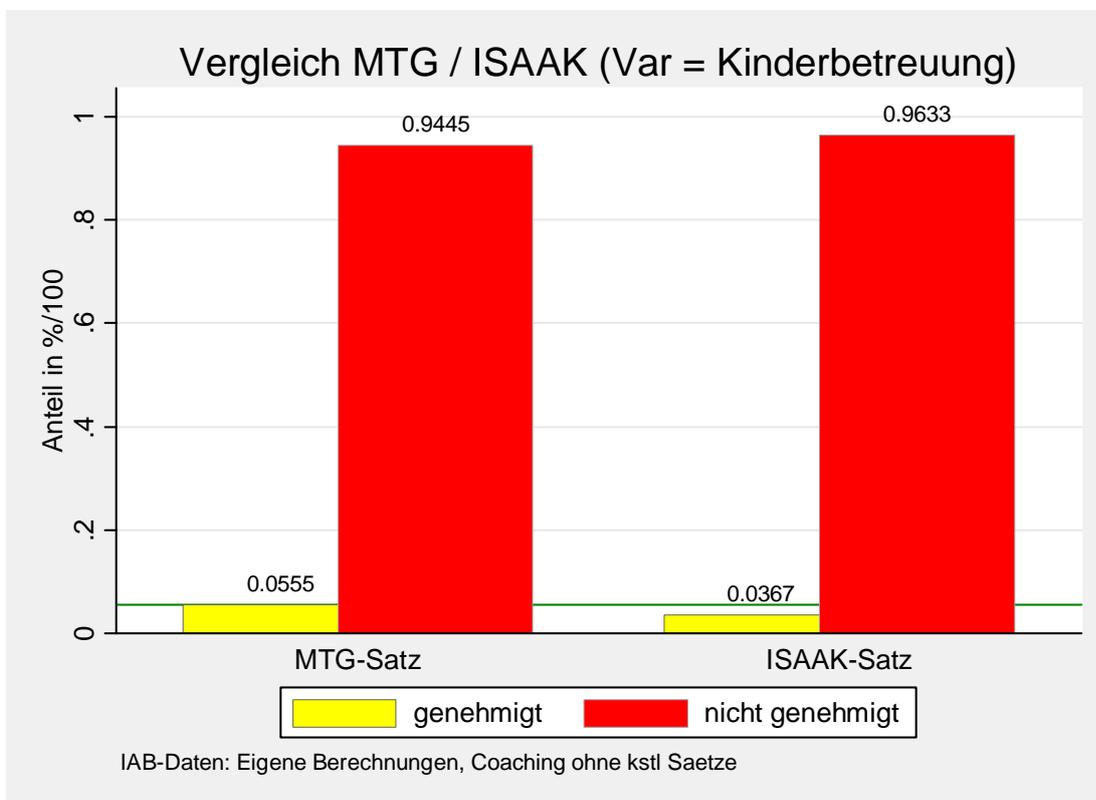
\*\*\*signifikant auf 99%-Niveau \*\* 95%-Niveau \* 90%-Niveau

x = getestet, - = nicht getestet

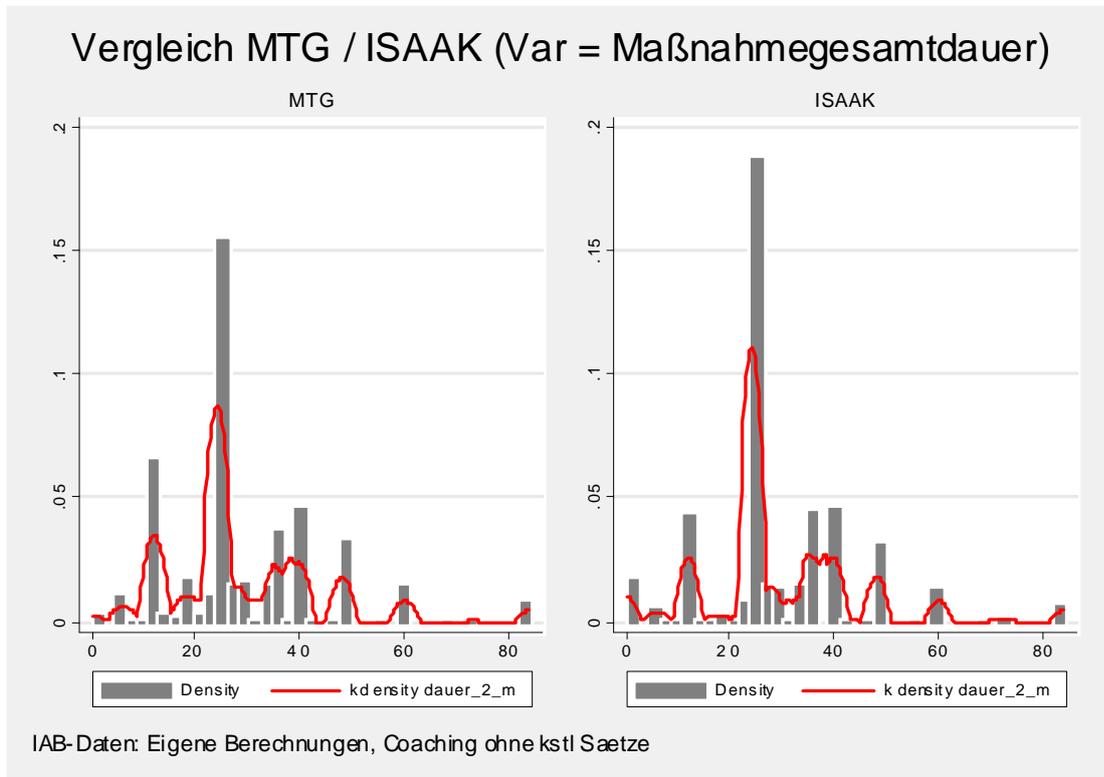
**Abbildung B3: Vergleich MTG / ISAAK; Anteil genehmigte und offene Fälle (Coaching)**



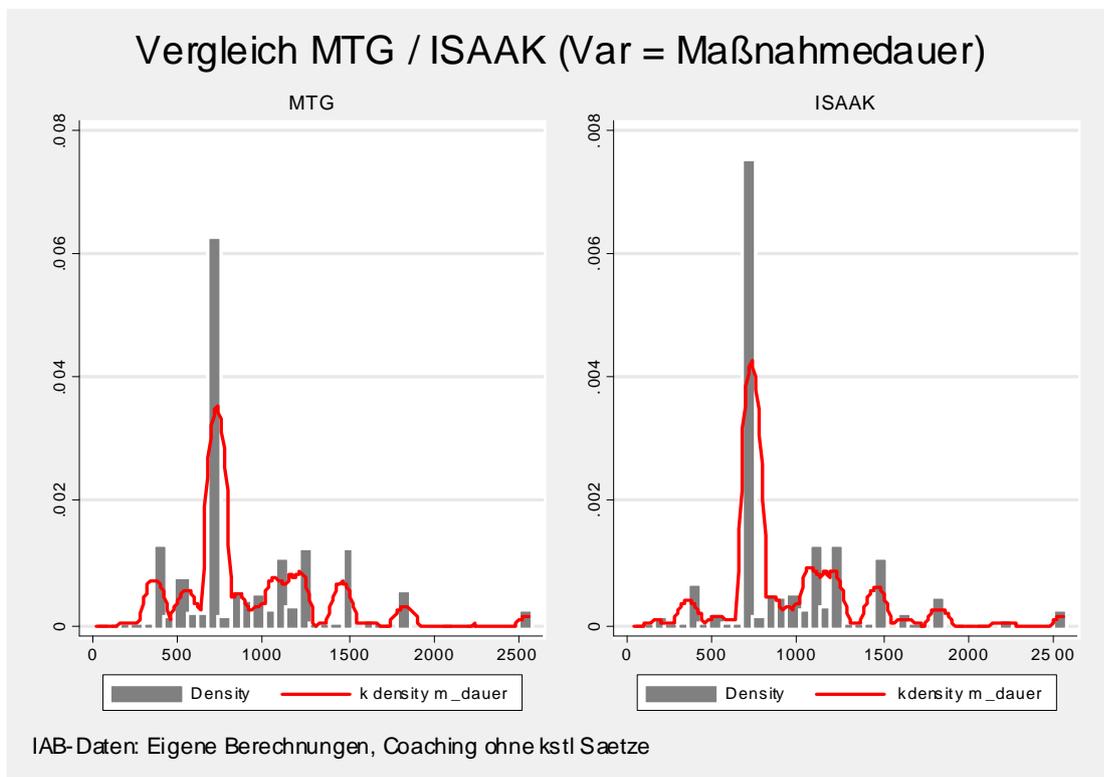
**Abbildung B4: Vergleich MTG / ISAAK; Anteil mit und ohne Kinderbetreuung (Coaching)**



**Abbildung B5: Vergleich MTG / ISAAK; Dichteverteilung Gesamtdauer (Coaching)**



**Abbildung B6 Vergleich MTG / ISAAK; Dichteverteilung Maßnahmedauer (FBW)**



### *Imprint*

#### **FDZ *Methodenreport***

No. 08/2008

#### **Publisher**

The Research Data Centre (FDZ)  
of the Federal Employment Service  
in the Institute for Employment Research  
Regensburger Str. 104  
D-90478 Nuremberg

#### **Editorial staff**

Stefan Bender, Dagmar Herrlinger

#### **Technical production**

Dagmar Herrlinger

#### **Copyright**

Reproduction – also in parts – only with permission of  
the FDZ

#### **Download**

[http://doku.iab.de/fdz/reporte/2008/MR\\_08-08.pdf](http://doku.iab.de/fdz/reporte/2008/MR_08-08.pdf)

#### **Internet**

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

#### **Corresponding author**

Dirk Oberschachtsiek, Institute for Employment  
Research, Regensburger Str. 104,  
D-90478 Nuremberg  
Phone: +49-(0)911/179-1615  
E-Mail: [dirk.oberschachtsiek@iab.de](mailto:dirk.oberschachtsiek@iab.de)