

Ein DataWarehouse für Forschung - Mit dem regionalen Online Analyse System des IAB ist ein erster wesentlicher Schritt getan

Jürgen Passenberger, Manuela Reith

1 Aufbruch zu Information excellence

1.1 Ausgangssituation

Das IAB als Abteilung VI der Bundesanstalt für Arbeit stützt seine Forschung auf eine Vielzahl empirischer Daten. Insbesondere greift das IAB - durch gesetzliche Vorgaben legitimiert - auf die breite Basis der Geschäftsdaten der BA sowie der Beschäftigtenstatistik zurück. Darüber hinaus werden eigenerhobene bzw. in Kooperation mit anderen Forschungsinstituten erhobene Daten sowie Daten anderer Institutionen (z. B. statistischer Ämter, Wirtschaftsforschungsinstitute) verarbeitet. Ziel ist es, aufbauend auf diesen Datenquellen für Arbeitsmarktpolitik und Arbeitsverwaltung fundierte Entscheidungshilfen bereitzustellen sowie Informationen aus der Arbeitsmarkt- und Berufsforschung für die Öffentlichkeit zugänglich zu machen.

Um den wachsenden Informationsbedarf aus Politik, Arbeitsverwaltung und Öffentlichkeit decken zu können, werden von der Arbeitsmarkt- und Berufsforschung immer schneller immer differenziertere Ergebnisse gefordert. Damit steigen auch die Ansprüche an Umfang und Qualität der Datenbasis sowie an die Fähigkeiten und Kapazitäten der Datenverarbeitung im IAB.

Auf den ersten Blick scheint die vorhandene, sehr ausdifferenzierte Arbeitsmarktstatistik eine hinreichende Grundlage für den Forschungsprozess zu bieten. Trotz der Datenvielfalt gibt es aber mit den bisherigen Systemen und Verfahren bei der Verdichtung und Verknüpfung der Daten sowie bei der schnellen Verfügbarkeit von konsistenten Zeitreihen erhebliche Defizite: Die Datenbestände der Bundesanstalt für Arbeit (BA) liegen derzeit auf unterschiedlichen Plattformen unterschiedlich strukturiert vor. Sie sind nicht miteinander verknüpft und in aller Regel nicht an den aktuellen Gebietsstand angepasst, d. h. regionsbezogene Datenwerte liegen für unterschiedliche Jahre zum jeweils in diesem Jahr geltenden Gebietsstand vor. Als Folge verbringen die Wissenschaftler sehr viel Zeit mit Suche und Aufbereitung von Daten für die weitere Analyse bzw. Präsentation.

1.2 Vision

Um diese Defizite zu beseitigen und eine deutliche Verringerung des Datenaufbereitungsaufwandes bei gleichzeitiger Steigerung der Analysemöglichkeiten und -qualität zu

erreichen, begann man im IAB Ende 1998 mit dem Aufbau des *plattformübergreifenden allgemeinen Analyse- und Informationssystems pallas*. Dieses soll den Wissenschaftlern im IAB nicht nur die Möglichkeit zur leistungsfähigen Analyse auch großer Datenmengen mit klassischen Werkzeugen der statistischen Analyse (z. B. SAS, SPSS) bieten, sondern darüber hinaus auch ein leistungsfähiges Online Analyse System. Verbunden mit *pallas* ist die Vision, mit Hilfe eines integrierten Analyse-, Informations- und Wissensmanagementsystems als zentrales Nervensystem der „Denkfabrik“ IAB, den Wissenschaftlern ein leistungsfähiges „Produktions- und Distributionssystem“ zu schaffen, das durch ineinandergreifende Prozesse und Werkzeuge den Forschungsprozess unter inhaltlichen, technischen und organisatorischen Aspekten weitestgehend integriert und optimiert und damit die Voraussetzung schafft, die Qualität und Effizienz der statistischer Analysen im IAB nachhaltig zu verbessern.

Den Kern dieser Vision von einem neuen innovativen System bildet ein leistungsfähiges Online Analyse System auf der Basis von DataWarehouse-Technologie¹. Durch Entwicklung einer qualitativ hochwertigen Forschungsdatenbasis, die eine flexible Verknüpfbarkeit von Daten aus unterschiedlichsten Quellen erlaubt, und durch den Einsatz von leistungsfähigen Werkzeugen für eine performante multidimensionale Analyse sollen die Forscher von aufwendigen Aufbereitungs- und Koordinierungsaufgaben entlastet werden und sich auf ihre Kernaufgabe - die Generierung fundierter entscheidungsrelevanter Informationen für Arbeitsmarktpolitik und Arbeitsverwaltung - konzentrieren können.

Darüber hinaus soll das System auch den BA-internen und -externen Kunden des IAB

- Politikern und Entscheidern in der Arbeitsmarktpolitik und Arbeitsverwaltung
- externen Wissenschaftlern
- Wirtschaftsorganisationen sowie
- der interessierten Öffentlichkeit

über Internet einen direkten flexiblen und leichten Zugang zu ausgewählten Datenbeständen und wichtigen Kennzahlen der Arbeitsmarkt- und Berufsforschung ermöglichen.

¹ Ein **DataWarehouse** bezeichnet eine themenorientierte, integrierte, zeitbezogene und dauerhafte Sammlung von Daten zur Entscheidungsunterstützung des Managements. Herkömmliche Management-Informationssysteme (MIS) und Decision-Support-Systeme (DSS) verarbeiten Daten in der Regel direkt aus operativen Systemen heraus in ihrer originären Struktur. Im Gegensatz dazu sind in einem DataWarehouse Daten aus unterschiedlichen Quellen in einer einheitlich strukturierten (homogenen und integrierten) Datenbasis zusammengeführt und werden über ein Datenbank-Managementsystem verwaltet. Aufbereitet und qualitätsgeprüft stehen sie für zielorientierte und effiziente Abfragen und Analysen unterschiedlicher Fachanwenderkreise bereit.

Leitgedanke der Vision ist dabei die Schaffung von „information excellence“ nach den Kriterien

- die richtige Information
- in der richtigen Form
- zum richtigen Zeitpunkt
- für den richtigen Anwender

als Basis für eine ausgezeichnete Qualität der Prozesse und Ergebnisse der Informationsgenerierung im IAB.

Die Zeit für ein solches Vorhaben war günstig, da parallel die Entwicklung eines BA-weiten DataWarehouses begann, aus dessen zentraler konsolidierter Datenbasis sich das IAB-System mit allen BA-Geschäftsdaten würde versorgen können, wodurch eine wesentliche Voraussetzung für das Gelingen dieses anspruchsvollen und schwierigen Vorhabens gegeben war.

1.3 Erste Etappe: das regionale Online Analyse System des IAB

Natürlich ist die Realisierung einer derart anspruchsvollen Vision ein ebenso komplexes wie zeit- und kostenintensives Vorhaben, das mehr Zeit braucht als ein oder zwei Jahre. Deswegen begannen wir im ersten Schritt zunächst mit der Entwicklung eines regionalen Online Analyse Systems als erster Stufe des geplanten DataWarehouse für die Forschung. Zu Beginn des Jahres 2002 - nach einem etwa zweijährigen Prototyping, das der Markterkundung und Klärung der technischen Machbarkeit diente, und daran anschließender rund einjähriger Entwicklungsdauer - wurde dieses System in einer ersten Version fertiggestellt und kann nun von den Regionalforschern im IAB und in den Landesarbeitsämtern am Arbeitsplatz für die Erstellung von Abfragen im Rahmen ihres Tagesgeschäfts eingesetzt und „auf Herz und Nieren“ geprüft werden.

Im ersten Realisierungsschritt des regionalen Online Analyse Systems hatten wir uns primär auf die Realisierung der Kernelemente des geplanten DataWarehouse-basierten Systems konzentriert:

- eine integrierte grundkonsolidierte Datenbasis, die wesentliche Datenquellen und Kennzahlen der Regionalforschung enthält, sowie
- performante und komfortable Funktionen zur flexiblen multidimensionalen Online Analyse für die Wissenschaftler.

Ergänzende Funktionen wie z. B. eine komfortable Navigations- und Recherche-Oberfläche, DataMining, ein Geographisches Analyse System (GIS) oder komfortable Schnittstellen werden erst in weiteren Schritten realisiert, wie auch noch bestehende Potenziale

zur Optimierung der Datenqualität erst in den weiteren Realisierungsschritten systematisch analysiert und ausgeschöpft werden können. Trotz dieser Einschränkungen wurde bereits mit dieser ersten Version eine deutliche Verbesserung der Analysemöglichkeiten und -qualität erreicht, die sich insbesondere auch in einer erheblichen Reduzierung des für die Datenbereitstellung und -aufbereitung erforderlichen Aufwands niederschlägt.

2 Wesentliche Verbesserungen der Datenqualität

Den wohl größten Beitrag zur Verbesserung der Analysemöglichkeiten und -qualität bildet die Schaffung einer integrierten, für regionale Fragestellungen optimierten Datenbasis, die in auswertungsbereiter Form zeitnah aktualisiert für Analysen bereitgehalten wird. In einem Wissenschaftsbetrieb, dessen zentrale Aufgaben in der Durchführung vielfältigster Projekte, die vielfach Einmaligkeitscharakter besitzen, sowie in der schnellen Beantwortung unterschiedlichster ad hoc-Fragen aus Politik, Verwaltung und Öffentlichkeit bestehen, kann dabei nur ein Teil der Bedarfe durch die Bereitstellung standardisierter Kennzahlen oder Standardauswertungen erfüllt werden. Ziel eines DataWarehouse für die Forschung muss es vielmehr vorrangig sein, die Datenbasis so zu gestalten, dass die Daten aus allen relevanten Quellen ad hoc flexibel in beliebigen Verknüpfungen, Verdichtungen und Merkmalskombinationen abgefragt werden können. Diese Abfrageergebnisse bilden vielfach die Basis für weitergehende Analysen mit spezialisierten Werkzeugen oder müssen in präsentabler Form aufbereitet werden können. Diese zentrale Anforderung, aus dem System vorrangig die für die Forschung relevanten Daten sehr flexibel selektieren und aufbereiten zu können, stellt hohe Anforderungen sowohl an die Datenbasis als auch an die Analyse-, Abfrage- bzw. Präsentationsfunktionen des Systems.

Neben der schnellen und flexiblen Zugänglichkeit und zeitoptimalen Verfügbarkeit der für die Forschungsprozesse erforderlichen Daten kommt vor allem auch der inhaltlichen Qualität der Daten eine herausragende Bedeutung zu. Die besten Analyse- und Präsentationsfunktionen nutzen nichts, wenn die verwendeten Daten fehlerhaft oder unvollständig sind. Die Erzeugung und Sicherung einer den Qualitätsanforderungen der Anwender genügenden Datenbasis ist deshalb eine zentrale Aufgabe bei der Entwicklung und Pflege bzw. bei der Bewertung solcher Systeme.

Ganzheitliche Modelle der Daten- bzw. Informationsqualität unterscheiden häufig verschiedene Dimensionen der Qualität. Stellvertretend für eine Reihe ähnlicher Konzepte sei hier das im DataWarehouse-Kontext häufig zitierte Modell von Larry English² „Total Data Quality Management“ genannt, das folgende Dimensionen unterscheidet:

² Larry P. English: Improving DataWarehouse and Business Information Quality; New York u. a. 1999.

- Quality of Data Definition (Definitionsqualität): z. B. klare, präzise, vollständige und einzigartige Definitionen; vollständige, korrekte Regeln; Konsistenz der Datennamen
- Quality of Data Content (Inhaltsqualität): Korrektheit, Vollständigkeit, Eindeutigkeit, Genauigkeit, Konsistenz, Aktualität
- Quality of Data Presentation (Präsentationsqualität): Zugänglichkeit, Verständlichkeit, Benutzbarkeit, zeitoptimale Verfügbarkeit.

Selbstverständlich muss die IAB-Forschungsdatenbasis in allen diesen Dimensionen eine den Anforderungen der Regionalforschung (in weiteren Schritten den Anforderungen der gesamten Arbeitsmarkt- und Berufsforschung) genügende Qualität sicherstellen. Zusätzliche Probleme ergeben sich darüber hinaus aus der Anforderung zur Integration historischer (häufig nur in aggregierter Form vorliegender) Datenbestände, wie sie zur Bildung langer Zeitreihen erforderlich sind. Bei der Erstellung der jetzt vorgelegten Version des regionalen Online Analyse Systems konzentrierten wir uns deshalb zunächst auf die Verbesserung der zeitoptimalen Verfügbarkeit und Zugänglichkeit sowie die Vollständigkeit und Konsistenz der bereitgestellten Daten. Bereits hier zeigt sich ein großer Qualitätssprung im Vergleich zur bisherigen Situation.

2.1 Vollständigkeit

Es ist leicht nachvollziehbar, dass ein statistisches Analyse- und Informationssystem nur dann für einen Forscher seinen vollen Nutzen entfaltet, wenn er dort alle Daten und Kennzahlen findet und verarbeiten kann, die er für seine Arbeit braucht. Jede Datenquelle, die nicht in das System integriert ist, jede Kennzahl, die nicht ad hoc aus dem System abgerufen werden kann, bedeutet einen Bruch in der direkten Verknüpfbarkeit, in der Qualität, aber auch in den Analysemöglichkeiten - und damit erhöhten Bearbeitungsaufwand. Entsprechend ist eine Basis-Forderung der Regionalforscher die Bereitstellung aller relevanten BA-internen und -externen Daten mit allen verfügbaren Merkmalen sowie aller relevanten regionalen Kennzahlen in der *pallas*-Datenbasis.

Um schnell und flexibel auch aktuelle Entwicklungen einbeziehen zu können, benötigt Forschung die Daten aus den unterschiedlichen Datenquellen zeitnah fortgeschrieben, wenn auch die Aktualisierungsperioden - Monate oder Quartale - hier (noch) verhältnismäßig lang sind. Zusätzlich benötigt die Regionalforschung - je nach Fragestellung und Methode - mehr oder weniger lange Zeitreihen, woraus sich die Anforderung ableitet, die Datenbasis so weit wie möglich und fachlich sinnvoll in die Vergangenheit zu verlängern.

Mit dem Online Analyse System wurde ein erster wesentlicher Schritt zur Realisierung dieser Anforderung getan, indem die für die Regionalforschung zentralen Arbeitslosen- und Beschäftigtendaten im Online Analyse System nicht nur aktuell, sondern auch für die

Vergangenheit bereitgestellt werden. Auf der Ebene der Arbeitsmarktverwaltung liegen die wichtigsten Eckwerte des Arbeitsmarktes (Vormeldungen) monatlich seit 1982, teilweise auch für weitere Strukturmerkmale wie Geschlecht, Nationalität und Alter vor. Arbeitslose und bedeutende Untergruppen wie Langzeitarbeitslose werden im Pilotsystem ab 1993 (Kreise) bzw. ab 1996 (Gemeinden) zusätzlich auf der Ebene der Kommunalverwaltung bereitgestellt, ab 2001 sind diese Daten monatlich verfügbar.

Daten zu sozialversicherungspflichtig Beschäftigten werden im Pilotsystem seit 1987 jährlich (Stichtag 30.06.) auf Geschäftsstellen- und Gemeindeebene bereitgestellt. Seit 1991 gilt dies nicht nur für den Arbeitsort, sondern auch für den Wohnort. Zusätzlich zu Arbeits- und Wohnort liegen die Beschäftigtendaten für eine Vielzahl weiterer Strukturmerkmale vor, die bei Abfragen beliebig kombiniert werden können. Gerade bei dieser Datenquelle mit ihrem großen Datenvolumen und der Komplexität der Merkmale wird der Mehrwert der eingesetzten MOLAP-Technologie besonders deutlich (siehe Abschnitt 3)

Darüber hinaus stehen als regionale Kennzahlen die amtliche Arbeitslosenquote und die Gebietsfläche zur Verfügung sowie alle Daten, die als Bezugsgrößen zur Berechnung von regionalen Arbeitslosen-³, Unterbeschäftigungs-⁴ und alters- und geschlechtsspezifischen Erwerbsquoten⁵ erforderlich sind (siehe Übersicht 1 in Abschnitt 2.2). Neben BA-eigenen Datenquellen wurden dazu auch externe Daten (z. B. des Statistischen Bundesamtes) gemäß den Benutzeranforderungen in die Datenbasis einbezogen.

Vollständigkeit bedeutet auch, dass die Daten lückenlos vorliegen. Für die regionale Arbeitsmarktforschung schlägt sich dies in der wichtigen Anforderung nieder, dass alle re-

³ Die **Arbeitslosenquote** ist die in der Öffentlichkeit wohl am häufigsten diskutierte Kenngröße für das Beschäftigungsdefizit. Sie setzt die Zahl der arbeitslos gemeldeten Personen in Relation zu den zivilen abhängigen Erwerbspersonen (sozialversicherungspflichtig Beschäftigte + geringfügig Beschäftigte + Beamte + Auspendler + Arbeitslose).

⁴ Da durch den unterschiedlichen Einsatz arbeitsmarktpolitischer Instrumente unterschiedlich hohe Arbeitslosenquoten hervorgerufen werden können, werden häufig auch Kenngrößen verwendet, die die arbeitsmarktpolitischen Aktivitäten „herausrechnen“. Besondere Bedeutung kommt hier der **Unterbeschäftigungsquote** zu, die die ungeforderte Beschäftigung (Arbeitslose + ABM-Teilnehmer + SAM-Teilnehmer + FBW-Teilnehmer + Vollzeitäquivalent der Kurzarbeiter) in Relation zu den zivilen abhängigen Erwerbspersonen (sozialversicherungspflichtig Beschäftigte + geringfügig Beschäftigte + Beamte + FbW-Teilnehmer + Auspendler + Arbeitslose) setzt. Naturgemäß ist die Unterbeschäftigungsquote höher als die Arbeitslosenquote.

⁵ Die **Erwerbsquote** ist ein Maß für die Erwerbsbeteiligung. Sie ist definiert als Quotient aus der Erwerbspersonenzahl (Arbeitslose + Erwerbstätige) und der Zahl der Wohnbevölkerung im Alter von 15 bis 64 Jahren). Die Erwerbstätigen wiederum setzen sich zusammen aus sozialversicherungspflichtig Beschäftigten + geringfügig Beschäftigten + Beamten + Selbständigen + mithelfenden Familienangehörigen + Auspendler.

levanten Daten für alle relevanten regionalen Einheiten zur Verfügung stehen. Ein wesentliches Problem im Rahmen regionaler Analysen ist jedoch (vor allem für die Vergangenheit), dass ein Teil der erforderlichen Daten ausschließlich in der Gebietsgliederung der Arbeitsverwaltung (Geschäftsstelle, Arbeitsamt, Landesarbeitsamt), ein anderer Teil ausschließlich in der Gebietsgliederung der Kommunalverwaltung (Gemeinde, Kreis, Regierungsbezirk, Bundesland), verfügbar ist. Leider sind diese beiden Gebietsgliederungen in den unteren Ebenen im Zuschnitt nicht deckungsgleich und auch nicht gegenseitig eindeutig zuordenbar. Für die Durchführung einer regionalen Analyse, in die Daten beider Regionalgliederungstypen einbezogen werden sollen, mussten insbesondere für historische Datenbestände, die nur in aggregierter Form vorliegen, bisher vom Forscher aufwendige Umschlüsselungs- und (in den Fällen, wo eine eindeutige Ableitung aus tieferen regionalen Einheiten nicht möglich ist) Umschätzprozeduren durchgeführt werden, um alle erforderlichen Daten zu einem einheitlichen Gebietsstand verfügbar zu haben.

Für das regionale Online Analyse System wurden in Zusammenarbeit mit den Regionalforschern Lösungen entwickelt, die es ermöglichen, die Datenbasis - in einem ersten Schritt ab Kreis- bzw. Arbeitsamtsebene - lückenlos sowohl in der Gebietsgliederung der Kommunal- als auch der Arbeitsverwaltung jeweils aktualisiert bereitzustellen, sodass der Forscher nun quasi „auf Knopfdruck“ online die Daten für die gewünschte regionale Gliederung abrufen kann. Wichtige Daten zum Arbeitsmarkt, die bisher nur für die Gebietsgliederung der Arbeitsverwaltung vorlagen, können damit nun zusätzlich auch nach der regionalen Gliederung der Kommunalverwaltung analysiert werden. Darüber hinaus ist die flexible Bezugsetzung der verschiedenen Daten und Kennzahlen sowohl auf Arbeitsamts- als auch auf Kreisebene möglich.

2.2 Konsistenz

Zweifellos wird die Bereitstellung und Sicherung einer konsistenten - im Sinne von widerspruchsfrei - integrierten Datenbasis einen wesentlichen Quantensprung für die Arbeitsmarkt- und Berufsforschung bedeuten. Gerade hier ist jedoch vor der Realisierung dieses Ziels eine Vielzahl von Problemen zu lösen, die zu großen Teilen weniger datentechnischer als vielmehr fachlicher Natur sind. Neben Fragen, die durch die Verknüpfung von Datenbeständen aus verschiedenen Quellen aufgeworfen werden, sind vor allem für die Integration der historischen Datenbestände gemeinsam mit den Forschern Konzepte zu entwickeln. Z. B. kollidiert die grundlegende Anforderung, dass es für jedes Datum und jede Kennzahl im System nur einen „richtigen“ Wert geben darf, mit einem weiteren ebenfalls grundlegenden Konsistenzkriterium, das sich aus der Anforderung der Forschung nach weit in die Vergangenheit reichenden Zeitreihen ableitet: keine Brüche zwischen den über das bisherige System zur Verfügung stehenden Daten und den Daten aus dem neuen System. Wenn es also in der Vergangenheit Inkonsistenzen in den bestehen-

den Statistiken gab, sollen diese nun bei Übernahme der historischen Daten zur Sicherung langer Zeitreihen im neuen System entsprechend abgebildet werden? Oder ist ein „richtiger“ Wert festzulegen, mit der Folge, dass bestimmte Zeitreihen eben nicht mehr ohne Bruch gebildet werden können. Es wäre zu viel erwartet, dass bereits in der ersten Version des regionalen Analyse Systems die im Zusammenhang mit der Konsistenz auftretenden Probleme vollständig oder zumindest zu großen Teilen hätten bereinigt werden können. Mit Hilfe der im System bereitgestellten integrierten Datenbasis sind jedoch Entwickler und Fachanwender im Rahmen der Testphase nun in der Lage systematisch Inkonsistenzen zu identifizieren und zu analysieren, um auf dieser Basis fachliche Lösungen zu entwickeln und zu bewerten.

Ein in diesem Zusammenhang zentrales Problem ist die Herstellung konsistenter Zeitreihen für Regionalanalysen: Stellen Sie sich vor, Sie betrachten die Arbeitslosenzahlen des Arbeitsamts Chemnitz und stellen fest, dass diese von Dezember 1995 auf Januar 1996 von 35.379 auf 44.676 steigen. Dies ist ein Zuwachs von rund 26 % in nur einem Monat. Ursache könnte z. B. die Schließung eines größeren Unternehmens oder eine größere Entlassungsaktion sein. Nicht immer jedoch sind Sprünge in einer Zeitreihe auf solche strukturellen Effekte zurückzuführen; mögliche Ursache können auch Veränderungen im Zuschnitt der betrachteten Gebietseinheit (in diesem Fall dem Arbeitsamt Chemnitz) sein. Tatsächlich waren ab Januar 1996 die Dienststellen Brand-Erbisdorf und Freiberg vom Arbeitsamt Pirna (dessen Arbeitslosenzahlen entsprechend sanken) zum Arbeitsamt Chemnitz gewechselt und haben ihre jeweiligen Arbeitslosenzahlen „mitgenommen“. Der zu beobachtende Anstieg der Arbeitslosenzahlen in Chemnitz war zu erheblichen Teilen auf diesen Effekt zurückzuführen. In der amtlichen Statistik werden derartige Veränderungen im Gebietsstand rückwirkend nicht betrachtet, da sie für deren Fragestellungen nicht relevant sind. Für statistische Analysen der regionalen Arbeitsmarktforschung dagegen stellen homogene und konsistente - um regionale Effekte bereinigte - Zeitreihen eine notwendige Basis dar. Voraussetzung hierfür ist die Konsolidierung der Datenbasis für einen einheitlichen (i. d. R. ist dies der aktuell gültige) Gebietsstand, wofür eine Anpassung des historischen an den aktuellen Gebietszuschnitt erforderlich ist.

Aufwendig, aber methodisch noch relativ einfach ist dies, wenn die Zuordnung von historischen zu aktuellen Gebietsständen anhand von kleineren Einheiten (i. d. R. über Gemeinden oder Geschäftsstellen) erfolgen kann. Hierzu müssen große Zuordnungsmatrizen aus einer Vielzahl heterogener Datenquellen, die sich auf viele verschiedene Zeitpunkte beziehen, aufgebaut werden. Alle jemals vorhandenen Gebietszuschnitte müssen über diese Matrizen dem aktuellen Gebietsstand zuordenbar sein.

Problematischer wird eine Anpassung an den aktuellen Gebietsstand dann, wenn Informationen aus kleineren Einheiten zur Zuordnung nicht zur Verfügung stehen (dies ist bei

nur aggregiert vorliegenden historischen Daten häufiger der Fall) oder wenn eine eindeutige Zuordnung anhand der kleineren regionalen Einheiten nicht möglich ist, z. B. weil sich die kleineren Einheiten selbst verändert haben oder weil die Änderung des Gebietszuschchnitts über bloße Zusammenlegungen hinausgeht. In solchen Fällen müssen Umschätzungen vorgenommen werden.

Der gesamte Prozess der Gebietsstands Anpassung ist außerordentlich aufwändig und erfordert große Sorgfalt, um keine übermäßigen Unschärfen zu produzieren, die das Analysepotenzial der Daten gefährden. Dies gilt insbesondere bei Regionalanalysen für Ostdeutschland, wo vor allem in den 90er Jahren nicht nur für Stadt- und Landkreise, sondern auch für die Dienststellen der Bundesanstalt für Arbeit der Zuschnitt der Regional-einheiten ständigen Veränderungen unterworfen war.

Im Pilotsystem wurden mit dem für die Regionalforschung zuständigen Fachbereich im Rahmen einer Entwicklungskooperation Zuordnungsmatrizen und Umschätzalgorithmen zur Konsolidierung der Zeitreihen entwickelt und auf die gesamte integrierte Datenbasis - wo fachlich erforderlich - angewandt:

- Verfahren zur Umschlüsselung historischer Gebietsstände von Gemeinden oder Geschäftsstellen auf den jeweils aktuellen Gebietsstand
- Verfahren zur Umschätzung historischer Gebietsstände von Kreisen oder Arbeitsämtern auf den jeweils aktuellen Gebietsstand.

Dadurch ist es möglich, alle auf Gemeinde- oder Dienststellenbasis vorliegenden Daten zusätzlich zum historischen Gebietsstand auch mit einem einheitlichen aktuellen Gebietsstand anzubieten, um so die Erzeugung konsistenter Zeitreihen sicherzustellen.

Zusammen mit der lückenlosen Bereitstellung aller Daten ab Kreis- bzw. Arbeitsamts-ebene sowohl für die Gebietsgliederung der Kommunal- als auch der Arbeitsverwaltung muss die automatische Aufbereitung aller Daten auf den jeweils aktuellen Gebietsstand als wesentlicher Beitrag für den fachlichen Mehrwert des Online Analyse Systems bezeichnet werden: Was bisher in sehr aufwendigen Aufbereitungsarbeiten durch den Wissenschaftler vorgenommen werden musste, kann jetzt quasi „auf Knopfdruck“ abgefragt werden.

**Übersicht 1: Datenbasis des regionalen Pilotsystems zur Online Analyse
- pallas.reg - Evolutionsstufe 1.1 (1. Quartal 2002)**

Daten	Regionale Basis-einheit	Strukturmerkmale	Periodizität	Zeitreihen	Quelle
Bestandszahlen					
Arbeitslose	Gemeinde; Arbeitsamt; (bis 12/94 West bzw. 12/95 Ost nur Kreis)	Geschlecht, Alter, Staatsangehörigkeit, Schwerbehinderung, Langzeitarbeitslosigkeit, vorher Angestelltenberuf, Arbeitszeit	quartalsweise monatlich (ab 01/01)	01/93 - aktueller Rand	Bundesanstalt für Arbeit: verschiedene Statistiken
Bevölkerung	Gemeinde; Arbeitsamt (teilweise Geschäftsstelle)	Geschlecht Altersgruppen (nur auf Kreisebene und bis 1999)	jährlich (31.12.)	1993 - 2000	Gemeindeverzeichnis des Statistischen Bundesamtes Fortgeschriebene Bevölkerungsdaten des Bundesamtes für Bauwesen und Raumordnung
Kurzarbeitende	Arbeitsamt	Wirtschaftszweig, Geschlecht, Arbeitsausfall	monatlich	01/87 - aktueller Rand	Bundesanstalt für Arbeit
Sozialversicherungspflichtig Beschäftigte	Gemeinde; Geschäftsstelle	Alle Strukturmerkmale der Beschäftigtenstatistik (flexibel kombinierbar)	jährlich (30.06.)	1980, 1987 - 2000	Bundesanstalt für Arbeit
Vormeldungen wichtiger Arbeitsmarktdaten					
Offene Stellen Vermittlungen Arbeitssuchende und Arbeitslose ABM/SAM-Teilnehmer Kurzarbeiter Beschäftigtenhilfe	Arbeitsamt	Geschlecht und je nach Kennzahl weitere Strukturmerkmale	monatlich	01/82 - aktueller Rand	Bundesanstalt für Arbeit
Kennzahlen					
Gebietsfläche	Gemeinde; (eingeschränkt Geschäftsstelle)	"---"	jährlich (31.12.)	1993 - 2000	
Amtliche Arbeitslosenquote	Kreis; Arbeitsamt (ab 1995)	ab 1995 Geschlecht, Alter, Staatsangehörigkeit	quartalsweise monatlich (ab 01/01)	IV/84 - aktueller Rand	Bundesanstalt für Arbeit

noch Übersicht 1

Daten	Regionale Basis-einheit	Strukturmerkmale	Periodizität	Zeitreihen	Quelle
Bezugsgrößen zur Berechnung von Quoten					
Arbeitslose Auspendelnde Grenz-arbeitnehmer Beamte Bevölkerung Geringfügig Beschäftigte Soz.vers.pfl. Beschäftigte Selbständige und mithelfende Familienangehörige	Kreis; Geschäftsstelle; Arbeitsamt	Geschlecht, Alter, Staatsangehörigkeit, Stellung im Beruf	jährlich, Monatswerte werden für amtl. vorgegebenen Zeitraum gesetzt	1993 - aktueller Rand	Bundesanstalt für Arbeit
ABM/SAM-Teilnehmende FbW-Teilnehmende Vollzeitäquivalent der Kurzarbeitenden	Kreis	"---"	quartalsweise (bis 12/00) monatlich (ab 01/01)	1995 - aktueller Rand SAM ab II/97 FbW ab IV/97	Bundesanstalt für Arbeit

3 Performante multidimensionale Analyse

Ziel des pallas Online Analyse Systems ist es, den Arbeitsmarktforschern und ihren Kunden die Möglichkeit zur performanten und flexiblen ad hoc-Analyse auf der gesamten integrierten Datenbasis zu schaffen. Eine leistungsfähige Möglichkeit hierzu bietet das sog. **Online Analytical Processing⁶ (OLAP)**, durch das die Fachanwender in die Lage

⁶ Online Analytical Processing (OLAP) steht für explorative, interaktive multidimensionale Analyse. Grundlage ist ein im DataWarehouse hinterlegtes multidimensionales Datenmodell. Datenanalyse mit OLAP ist ein dynamischer Prozess, bei der der Anwender mit Hilfe von multidimensionalen Operatoren durch die multidimensionale Datenstruktur navigiert:

- Pivotierung/Rotation: Diese Operation dreht den Würfel durch vertauschen der Dimensionen um seine Achsen und ermöglicht es dem Anwender, die Daten aus beliebiger Perspektive zu betrachten.
- Roll-up, Drill-down: Das Roll-up entspricht einer Verdichtung bzw. Aggregation der Daten innerhalb einer multidimensionalen Matrix, z. B. indem sich der Anwender z. B. von Tag über Monat und Quartal zu Jahr bewegt. Drill-down ist die dazu komplementäre Aktion.
- Drill-Across: Hier wechselt der Anwender von einem multidimensionalen Würfel (z. B. Arbeitslose) zu einem anderen (z. B. Beschäftigte).
- Slice und Dice: Hier werden aus den multidimensionalen Würfeln einzelne Scheiben herausgeschnitten.

versetzt werden, schnell und komfortabel multidimensionale ad hoc-Analysen durchzuführen, zu speichern und an Standard-Softwareprodukte wie Excel, Word und statistische Analysesysteme zu übergeben. Ergänzt werden soll die OLAP-Funktionalität durch ein Geographisches Informationssystem (GIS) und DataMining-Funktionalitäten.

Der Einsatz innovativer Web-Technologie soll zudem die zeitnahe dynamische Informationsbereitstellung in Intranet und Internet ermöglichen. Durch ein Portal mit leistungsfähigen Navigations- und Recherche-Funktionalitäten soll auch Systemnutzern ohne vertiefte Kenntnisse der Datenstrukturen in verständlicher Form ein einfacher und sicherer Zugang zu den Datenbeständen geboten werden.

Im regionalen Teilsystem ist bereits die volle OLAP-Funktionalität realisiert. Um die hohen Anforderungen der Arbeitsmarktforscher an Abfragegeschwindigkeit und -flexibilität erfüllen zu können, entschieden wir uns aufgrund der Erfahrungen im Prototyping zunächst für den Einsatz von OLAP auf der Basis multidimensionaler Datenbanktechnologie (MOLAP)⁷. Diese Form des Online Analytical Processing ist am besten in der Lage, die in der konsolidierten Datenbasis verfügbaren Daten und Kennzahlen über so performante und differenzierte ad hoc-Abfragen und -Vergleiche online zu präsentieren, wie dies in der Forschung erforderlich ist. Beispielsweise hat der Forscher Kennzahlen wie Arbeitslosenquoten auf der Ebene der Kreise nach Alter und Geschlecht differenziert nach verschiedenen Zeitpunkten oder Beschäftigtenzahlen nach Wirtschaftszweig, Berufsbildung und Arbeitszeit für Arbeitsamtsbezirke in Sekundenschnelle auf dem Bildschirm und kann dann die ausgewählten Daten weiter online vergleichen, verdichten oder selektieren. Dabei kann der Forscher in der integrierten Datenbasis über verschiedene Detaillierungsebenen hinweg navigieren (Drilldown-, Drillup-, Drillacross-Funktionalität). Beliebige Perspektiven auf das Datenmaterial sind möglich und bewirken einen Qualitätssprung in der Analyse arbeitsmarktrelevanter Fragestellungen.

Die erweiterte Leistungsfähigkeit durch MOLAP-Technologie lässt sich konkret am Beispiel der Beschäftigtendaten, die die Datenquelle mit dem bisher größten Volumen bil-

⁷ Je nach Art der Abbildung und Speicherung des multidimensionalen Datenmodells im DataWarehouse wird unterschieden zwischen

MOLAP: direkte Umsetzung des multidimensionalen Modells in multidimensionalen Datenbanksystemen (MDBMS)

ROLAP: Repräsentation des Modells auf einem relationalen Datenbankmodell

HOLAP (hybrides OLAP): Aufteilung der Umsetzung auf eine relationale und eine multidimensionale Datenbank.

(Siehe z. B. Bauer, Andreas/Günzel, Holger (Hrsg.) (2001): Data-Warehouse-Systeme: Architektur, Entwicklung, Anwendung, Heidelberg, S. 104 ff. sowie S. 197 ff.)

Der im IAB eingesetzte OLAP-Server unterstützt alle drei Arten des Online Analytical Processing.

den, darstellen.⁸ Alle im Pilotsystem zu dieser Datenquelle verfügbaren Merkmale waren grundsätzlich bereits auch über das bisherige System verfügbar, jedoch waren umfangreiche Verschachtelungen dieser Merkmale nur bedingt möglich und auch die Ausgabemöglichkeiten der Abfrageergebnisse waren begrenzt (z. B. durch die Beschränkung auf maximal 999 auszugebende Zeilen). Wollte man eine Auswertung für alle ca. 15.700 aktuell gültigen Gemeinden durchführen, konnte dies im bisherigen System nicht in einer Abfrage erfolgen, sondern musste auf mehrere Teilabfragen aufgeteilt werden. Das Pilotsystem bietet demgegenüber die Möglichkeit, alle vorhandenen Stichtage in einer Auswertung flexibel in den Zeilen oder Spalten auszugeben, wodurch sich die Bearbeitungszeiten der Datengewinnung im Vergleich zum bisherigen System drastisch reduzieren. Zum Beispiel dauert mit dem Pilotsystem die Generierung einer Abfrage zur Anzahl der Beschäftigten für alle Kreise (Zeilen) und die Jahre 1990 bis 2000 (Spalten) ca. drei Minuten, das sind rund 10 % der Bearbeitungszeit mit dem bisherigen System.

Außerdem ist es im Pilotsystem nun bei relativ kurzen Abfragezeiten möglich, beliebig viele der verfügbaren Merkmale - soweit fachlich sinnvoll - flexibel auf verschiedenen Aggregationsebenen in einer Abfrage miteinander zu kombinieren: so benötigt z. B. das Generieren einer Tabelle für das Jahr 2000, die die Zahlen der weiblichen Beschäftigten nach Ihrem Beruf (337 Ausprägungen) differenziert nach Alter (5-Jahresschritte) in den Zeilen sowie nach Stellung im Beruf differenziert nach Nationalität in den Spalten ausweist, eine Abfragezeit von weniger als 1 Minute.

⁸ Multidimensionales OLAP zeichnet sich durch analytische Mächtigkeit (sehr hohe Anfragegeschwindigkeit) und intuitive Bedienbarkeit aus, bei Würfeln steigender Größe, vor allem bei Dünnbesetztheit, erreicht man jedoch eine Skalierbarkeitsgrenze (siehe z. B. Bauer/Günzel Heidelberg 2001: 197 ff.). Aus diesem Grunde wird der Einsatz von MOLAP-Technologie vor allem für kleinere, bereits durch aggregierte Werte gebildete Datenbasen empfohlen, während bei großen Datenbeständen mit komplexen Hierarchien eher relationales OLAP empfohlen wird. Der Einsatz der Microsoft-Produkte MS SQL Server 2000 und MS Analysis Server mit ihrer intelligenten Speichertechnologie und hohen Skalierbarkeit ermöglicht es, die Vorteile des multidimensionalen OLAPs auch bei großen Datenmengen mit vielen Dimensionen einzusetzen. Beispiel: Der Bestand der sozialversicherungspflichtig Beschäftigten (jeweils zum 30.06.) liegt im Pilotsystem für 1980 und dann als Zahlenreihe ab 1987 vor. Pro Jahr gibt es ca. 27 Mio. Datensätze; der gesamte Datenwürfel hat ein Volumen von mehr als 14 GByte. Mit seinen 18 Dimensionen mit insgesamt 27 Dimensionshierarchien ergibt sich für den Datenwürfel pro Jahr die Anzahl von 1068 potenziellen Zellen. In der konventionellen MOLAP-Technologie würde jede dieser potenziellen Zellen, auch wenn sie leer ist, Speicherplatz beanspruchen, was (unterstellt man einen Speicherplatzbedarf von 1 Bit pro Zelle) eine Mindestgröße des Würfels von 10^{68} Bit = 10^{55} Tbyte (das ist eine 1 mit 55 Nullen!) bedeuten würde. Kein Rechner könnte einen Datenwürfel dieser Größe speichern. Durch Partitionierung und Eliminierung leerer Zellen reduziert sich die Größe des Würfels im MS Analysis Server tatsächlich auf lediglich - noch bequem zu managende - 1,5 GByte pro Jahr .

Zur professionellen Präsentation der OLAP-Abfragen bietet das Pilotsystem eine in das OLAP-Tool integrierte Möglichkeit zur Kartendarstellung sowie Darstellungsmöglichkeiten der OLAP-Abfrageergebnisse mit MS PowerPoint

- als statische Folien oder
- mit Drill-Möglichkeit in der PowerPoint-Grafik (Integration von ProClarity als OLE-Objekt in eine PowerPoint-Folie).

Für den Export nach MS Excel existiert ebenfalls eine - allerdings noch zu optimierende - Schnittstelle. Ergänzt wird die Werkzeugpalette des Pilotsystems durch die prototypische Realisierung einer Navigationsoberfläche, die einen thematischen Einstieg in die Online Analyse auch ohne vertiefte Kenntnisse der Datenbankstrukturen ermöglichen soll.

4 Dokumentation

Die Zugänglichkeit und Verwertbarkeit von Daten ist ohne klare, präzise und vollständige Beschreibung stark eingeschränkt; erst durch zusätzliche und verständliche Angaben über den Entstehungskontext, die inhaltliche Qualität und den Bedeutungsgehalt werden aus Daten für den Nutzer Informationen. In der Forschung sind die Ansprüche hier - notwendigerweise - besonders hoch, wie auch die Vielzahl von Fußnoten in Forschungsberichten zeigt, die nicht nur die Funktion haben, die Belesenheit des Autors zu belegen, sondern dem Leser die für eine weitere qualifizierte Verwendung erforderlichen Zusatz- bzw. Hintergrundinformationen liefern sollen. Entsprechend besteht ein hoher Anspruch an eine vollständige, verständliche und leicht zugängliche Dokumentationen zu den bereitgestellten Daten.

Um diesen Anspruch zu decken, plant das IAB mit dem Aufbau einer Datenbank für Fach-Metadaten der Arbeitsmarkt- und Berufsforschung Neuland zu betreten. Die in einer zentralen Metadatenbank hinterlegten und mit den Sachdaten verknüpften Datenbeschreibungen und fachlichen Zusatzinformationen sollen dem Informationssuchenden Hilfestellung leisten: beim Auffinden geeigneter Informationen ebenso wie beim Verstehen und Bewerten der gefundenen Ergebnisse. Damit soll ein wesentlicher Schritt zur Verbesserung der Informationsqualität und zum Aufbau einer personen-unabhängigen Wissensbasis im IAB getan werden. Die Beschaffung der hierfür erforderlichen Werkzeuge ist allerdings noch nicht abgeschlossen, sodass die Realisierung dieser Komponente noch zurückgestellt ist.

Einen ersten Beitrag in diese Richtung leistet das Pilotsystem - allerdings noch ohne Werkzeug-Unterstützung - bereits: Für die in die Datenbasis eingestellten Daten und Kennzahlen werden nicht nur Definitionen in für den Fachanwender verständlicher und leicht zugänglicher Form bereitgestellt werden, sondern auch Angaben zur Datenqualität

soweit hierzu aus den Qualitätsprüfungen des Entwicklungsteams bzw. der Testuser entsprechende Informationen zur Verfügung stehen.

5 Fazit: ein wichtiger Meilenstein ist erreicht

Mit der fertiggestellten ersten Version des regionalen Online Analyse Systems ist es gelungen, eine mit modernen leistungsfähigen Analysefunktionen auswertbare Datenbasis vorzulegen, wie sie der regionalen Arbeitsmarktforschung bisher nicht zur Verfügung stand:

- zentrale Daten und Kennzahlen der regionalen Arbeitsmarktforschung,
- integriert in einer grundkonsolidierten Datenbasis,
- in beliebiger Verdichtung und Merkmalskombination,
- in langen hinsichtlich der Gebietsdimension konsolidierten Zeitreihen,
- regelmäßig zeitnah aktualisiert,
- mit einer Grunddokumentation ergänzt und
- durch flexible ad hoc-Analysen leicht (auf „Knopfdruck“) auswertbar.

Ein wesentlicher fachlicher Mehrwert des Pilotsystems liegt damit in der erheblichen Verbesserung der Präsentationsqualität (Zugänglichkeit, zeitoptimale Verfügbarkeit) der für die Regionalforschung benötigten Daten. Bei Abfragen, die mit dem Pilotsystem statt mit dem bisherigen System durchgeführt wurden, zeigte sich, dass durch die Kombination von Vorhaltung auswertungsbereiter Daten und leistungsfähigen Analysewerkzeugen die Datenbereitstellungszeiten um bis zu 90 % reduziert werden konnten. Natürlich muss man davon ausgehen, dass bei einem Teil der Forschungsdatenbedarfe auch weiterhin aufwändige Sonderaufbereitungen erforderlich sein werden. Auch werden für eine verlässliche Abschätzung erst systematische Vergleiche zwischen Pilot- und Altsystem durchgeführt werden müssen. Dennoch geben die bisherigen Erfahrungen Anlass zu der Prognose, dass mit dem neuen System zukünftig erhebliche Verkürzungen der Bearbeitungszeiten bei der Datenbereitstellung erzielt werden können.

Für die inhaltliche und definatorische Datenqualität kann konstatiert werden, dass mit dem Pilotsystem das Ziel der ersten Etappe - die Realisierung einer Datenqualität, die mindestens der des Altsystems entspricht, - ohne Einschränkungen erreicht wurde. Damit ist sichergestellt, dass das den Regionalforschern jetzt zur Anwendung am Arbeitsplatz bereitgestellte System kein „Spielsystem“ ist, sondern von den Wissenschaftlern für die Bearbeitung ihrer Tagesaufgaben verwendet werden kann.

Sicherlich muss die endgültige Beurteilung des Ergebnisses wie auch des Prozesses der bisherigen Entwicklungsarbeiten letztendlich durch die Fachanwender erfolgen. Dennoch

bildet der Zwischenstand nach der ersten Etappe eine gute Basis für die weiteren Schritte zur Realisierung der Vision eines DataWarehouses bzw. Knowledge Warehouses für die Arbeitsmarkt- und Berufsforschung. Bis in ca. drei bis vier Jahren unsere Vision Wirklichkeit geworden ist, müssen allerdings noch einige wesentliche Schritte getan werden:

- Durchführung einer ausführlichen Test- und Optimierungsphase, in der gemeinsam mit den Fachanwendern die Datenqualität in ihren verschiedenen Dimensionen systematisch analysiert und fachliche Entscheidungen und Konzepte zur Qualitätssicherung und -verbesserung getroffen und getestet werden
- Ausbau der integrierten Datenbasis für die Regionalforschung und gleichzeitig Erweiterung um Bedarfe weiterer Fachbereiche des IAB, insbesondere der Evaluationsforschung (letzteres wird dadurch erleichtert und beschleunigt, dass wesentliche grundlegende Modellierungs- und Konsolidierungsarbeiten für die Regionalforschung auch für die übrigen Fachbereiche Gültigkeit haben)
- Systematische kontinuierliche Verbesserung der Datenqualität in allen Dimensionen durch Aufbau eines ganzheitlichen Systems zum Daten-/Informationsqualitätsmanagement
- Portierung des Pilotsystems auf die gemeinsam mit den BA DataWarehouse beschafften Tools
- Aufbau eines mit den Sachdaten verknüpften zentralen Metadatenmanagementsystems sowie
- Erweiterung der Werkzeug-Palette um weitere Analyse-Produkte wie z. B. Reporting oder DataMining
- Einbindung der BA-internen und -externen Kunden über Web-Technologie.

Der wesentliche Aufwand bei diesen geplanten Arbeiten (80-90 %) wird - dies haben die Erfahrungen aus den bisherigen Projektarbeiten gezeigt - in der Weiterentwicklung und Optimierung der integrierten auswertungsbereiten Datenbasis liegen, damit sie hinsichtlich ihrer Qualität den Bedarfen der Forschung genügt. Ausschlaggebender Erfolgsfaktor - auch dies haben die bisherigen Arbeiten deutlich gemacht - wird sein, dass es gelingt, die Fachbereiche konsequent und kontinuierlich aktiv in die Entwicklungsprozesse einzubinden. Hier bestehen sicherlich noch erhebliche Verbesserungspotenziale.

6 Und der Nutzen für den Arbeitslosen oder den Arbeitsmarkt?

In der Schlussdiskussion eines Vortrags über pallas stellte ein Zuhörer vor einiger Zeit die Frage: „Glauben Sie tatsächlich, dass Sie mit Ihrem System, so leistungsfähig es auch sein mag, etwas bewirken? Meinen Sie, dass dadurch die Anzahl der Arbeitslosen gesenkt oder die Arbeitsmarktpolitik effizienter wird?“ Meine Antwort auf diese Gewis-

sensfrage, die die entscheidende Frage überhaupt ist, war damals und ist auch weiterhin ein überzeugtes „Ja“. Natürlich hängt der Erfolg von Arbeitsmarkt- und Beschäftigungspolitik wie auch der Arbeitsverwaltung von vielen Faktoren ab - nicht zuletzt von der Kompetenz und der Verantwortungsbereitschaft der unterschiedlichen Akteure. Und natürlich ist ein System wie pallas nur ein Werkzeug und damit in seiner Wirksamkeit abhängig vom Gestaltungswillen und Sachverstand derjenigen, die dieses Werkzeug führen. Andererseits ist erfolgreiche Arbeitsmarktpolitik und Arbeitsverwaltung - sofern sie kein Zufallsprodukt sein soll - nicht möglich ohne eine rechtzeitig vorliegende fundierte und differenzierte Informationsbasis. Hierzu ist ein wesentlicher Quantensprung erreicht, wenn durch pallas die unschätzbaren Erkenntnispotenziale, die in den Geschäftsdaten der BA und den Beschäftigtendaten liegen, für das IAB wie auch die BA-externe Arbeitsmarktforschung in Form einer flexibel und performant auswertbaren Datenbasis erschlossen sein werden. Deswegen stellt der Aufbau von pallas als ein alle Möglichkeiten der innovativen Informationstechnologie ausschöpfendes System zur optimierten Datenbereitstellung und -analyse keine hinreichende, aber eine notwendige Voraussetzung dafür dar, dass die Arbeitsmarktforschung ihrer Verantwortung nachkommen und Politik und Verwaltung mit den richtigen Informationen in der richtigen Form und zur richtigen Zeit versorgen kann.

