

Dr. Claudia Fleck

GENESIS-Online

Die Internet-Datenbank des Statistischen Bundesamtes

In der heutigen Informationsgesellschaft spielt das Internet als Verbreitungsweg eine zentrale Rolle. Um den wachsenden Bedarf von Wirtschaft, Politik und Gesellschaft an statistischen Informationen zu erfüllen, bieten die Internet- und Datenbanktechnologien neue Möglichkeiten: So können statistische Informationen auf elektronischem Wege effizient und schnell gesammelt und verbreitet werden.

Mit dem Einsatz moderner Kommunikations- und Informationstechnologie unterstützt das Statistische Bundesamt als einer der größten Informationsdienstleister in Deutschland Produktion, Management und Verbreitung von Informationen. Im Rahmen der deutschen e-Government-Initiative wird das Statistische Bundesamt bis 2005 alle seine internetfähigen Dienstleistungen online bereitstellen. Mit der Freischaltung des statistischen Informationssystems „GENESIS-Online“ im Juli 2002 leistete das Statistische Bundesamt einen weiteren Beitrag zur „BundOnline 2005“-Initiative der Bundesregierung und verbesserte damit die Dienstleistungsqualität im Bereich der Informationsbereitstellung.

GENESIS heißt Gemeinsames Neues Statistisches Informationssystem – eine Datenbank, die von den Statistischen Ämtern des Bundes und der Länder entwickelt wurde, um ein gemeinsames Data-Warehouse für statistische Informationen aufzubauen.¹⁾ Über das Informationssystem besteht abhängig vom jeweiligen Betreiber des Systems ein Zugang zu Daten in Form von Aggregaten oder auch in tiefer regionaler bzw. sachlicher Untergliederung. GENESIS zeichnet sich u. a. durch eine Datenhaltung mit vollständiger Beschreibung der Inhalte über Metadaten mit der Möglich-

keit der metadatengestützten Recherche sowie durch eine komfortable Tabellenerstellung aus. Mit GENESIS-Online wurde vom Statistischen Bundesamt ein webbasierter Zugang zu GENESIS entwickelt, der es ermöglicht, das Statistische Informationssystem per Internet zu nutzen.

Im Folgenden werden das zugrunde liegende Datenmodell erläutert und die Funktionen von GENESIS-Online ausführlich vorgestellt. Im Abschnitt Systemarchitektur wird auf wesentliche technische Aspekte eingegangen. Anschließend werden das Datenangebot, das Marketingmodell mit den daraus resultierenden Zugangswegen sowie einige Eckzahlen zur Nutzung von GENESIS-Online betrachtet.

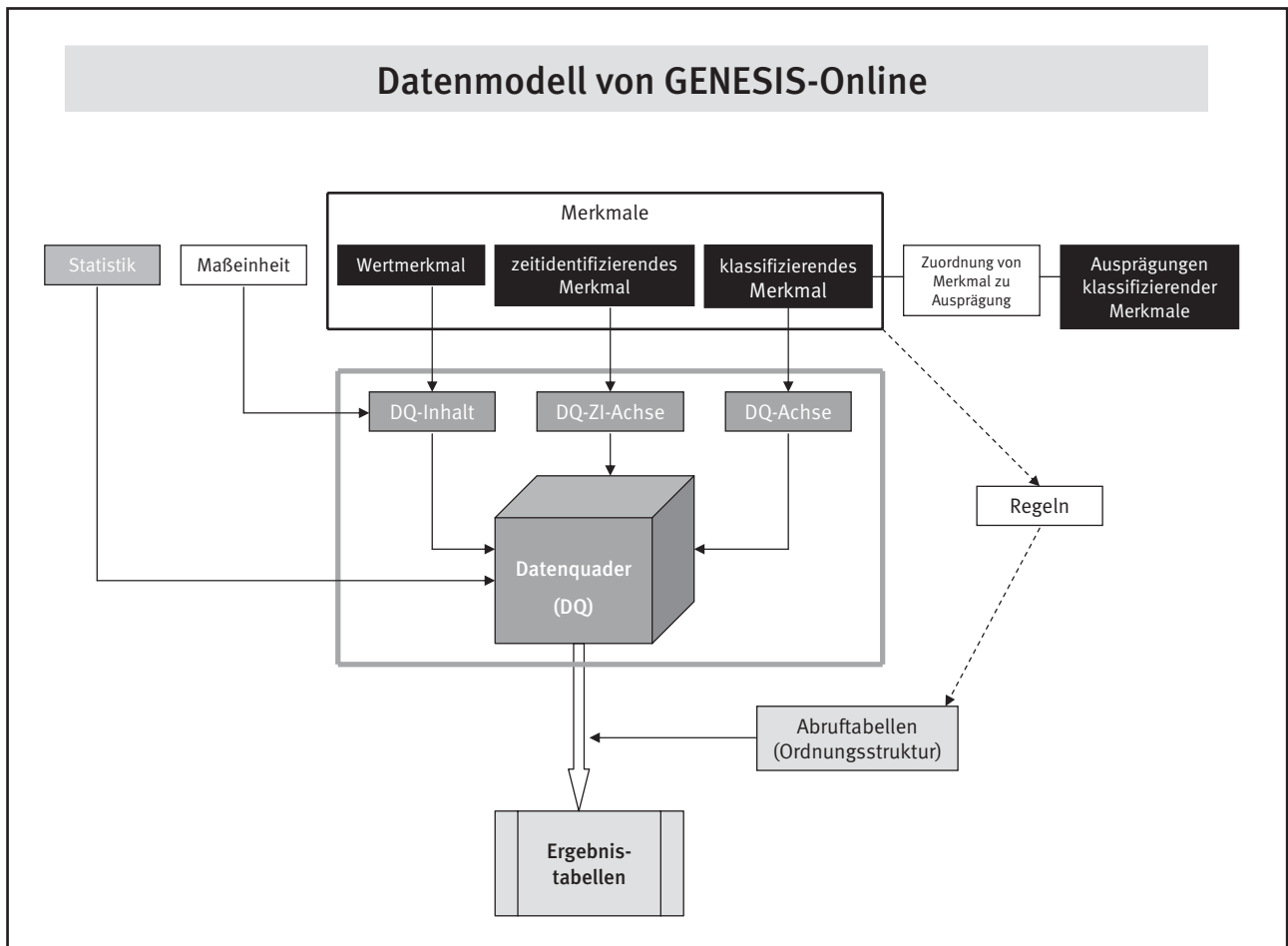
Das Datenmodell

In der amtlichen Statistik werden Daten für bestimmte Zeitpunkte oder Zeiträume erhoben, die durch festgelegte sachliche und räumliche Eigenschaften abgegrenzt werden. Der sachliche Hintergrund und die inhaltliche Aussage einer Zahl werden durch Metadaten beschrieben. GENESIS-Online enthält ein Referenzsystem für Metadaten mit Angaben zu der zugrunde liegenden Statistik, den erhobenen Merkmalen und ihren Merkmalsausprägungen und den Maßeinheiten, in denen die Werte nachgewiesen werden. Über Regeln können zudem neue Merkmale aus vorhandenen Merkmalen abgeleitet werden.

Das Datenmodell von GENESIS-Online ist das so genannte „Quadermodell“, das heißt die Werte werden in multidimen-

¹⁾ Siehe hierzu auch Rahm, H./Zipse, C.: „Konzeption und Entwicklung von GENESIS“ in dieser Ausgabe, S. 1092 ff.

Schaubild 1



sionalen Datenobjekten (Quadern) gespeichert, die durch die Metadaten beschrieben werden: Jeder Quader gehört zu genau einer Statistik und wird definiert durch Merkmale, die den Inhalt, die Zeit, die Region und andere Sachklassifikationen festlegen (siehe Schaubild 1). Der Inhalt eines Datenquaders wird durch das Wertmerkmal definiert. Das zeitidentifizierende Merkmal und die klassifizierenden Merkmale definieren die Achsen des Datenquaders und damit seinen zeitlichen, räumlichen und sachlichen Bezug. So wird zum Beispiel in der Statistik „Fortschreibung des Bevölkerungsstandes“ das Wertmerkmal „Bevölkerungsstand“ durch die Merkmalsausprägungen der klassifizierenden Merkmale „Bundesländer“, „Geschlecht“ und „Altersjahre“ räumlich und sachlich gegliedert. Das zeitidentifizierende Merkmal ist in diesem Fall der „Stichtag“ mit den Ausprägungen 31.12. der einzelnen Berichtsjahre. In einem Datenquader können auch mehrere Wertmerkmale gespeichert werden, sofern ihre Gliederung über die Datenquader-Achsen identisch ist. Die Verbindung der Werte mit den Metadaten ist die Grundlage für die Erschließung der Datenbestände.

Zur Tabellierung der statistischen Werte wird über die Anordnung von Metadaten in einer so genannten „Abruftabelle“ der Aufbau einer gewünschten Tabelle definiert. Eine Abruftabelle ist also eine Ordnungsstruktur, in der die Inhalte der einzelnen Strukturelemente einer Tabelle festge-

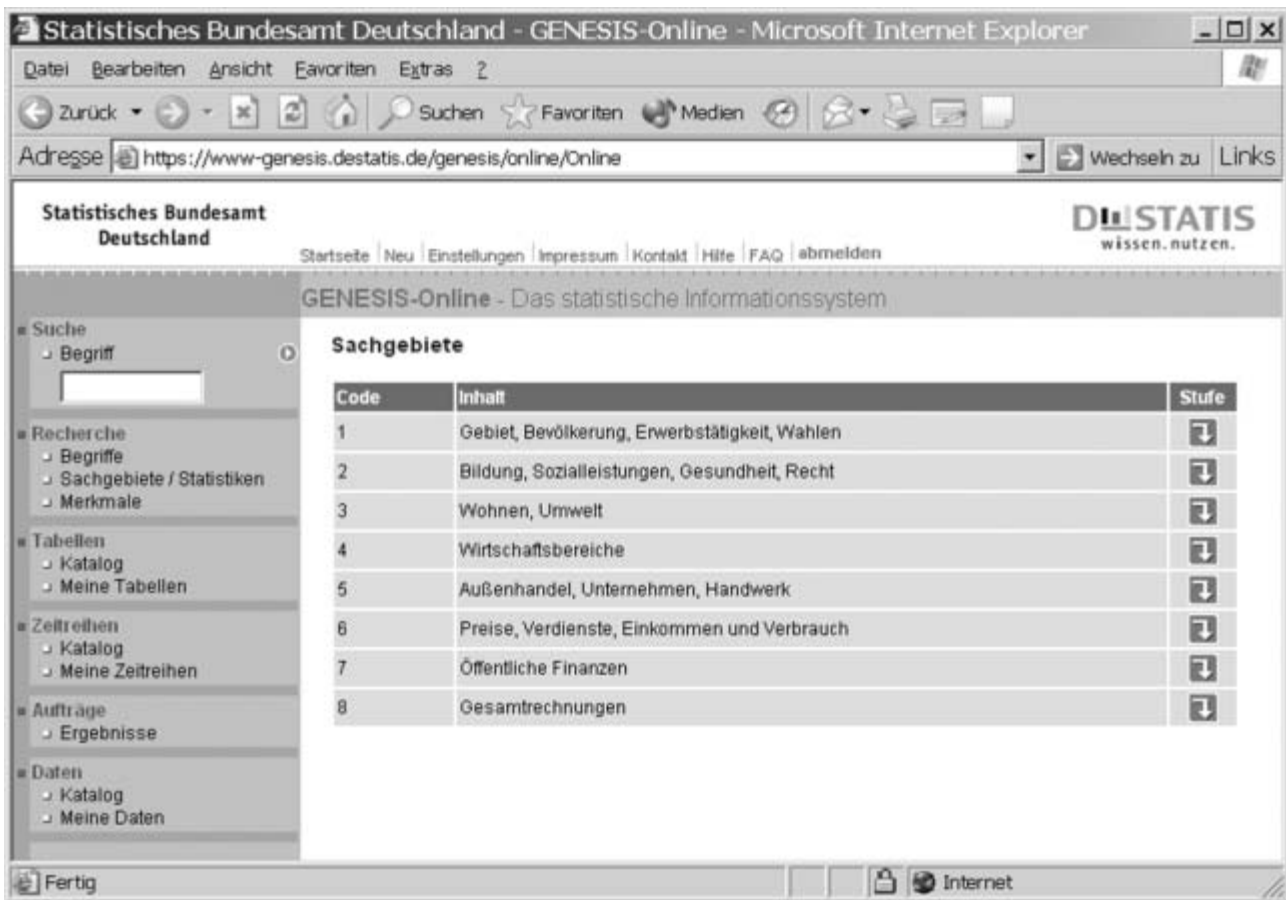
legt werden. Strukturelemente einer Tabelle sind Kopfzeile, Vorspalte, Untertitel, Zwischentitel sowie globale Angaben, die für die ganze Tabelle gültig sind, wie zum Beispiel die Statistik. Über den Werteabruf wird dann entsprechend den Vorgaben der Abruftabelle eine Ergebnistabelle erzeugt, das heißt der über die Abruftabelle definierte Tabellenrahmen wird mit Werten gefüllt, die in einem oder auch in mehreren Datenquadern gespeichert sind.

Das Quadermodell ermöglicht es, sowohl sachlich tief gegliederte Strukturdaten als auch Regionaldaten, Verflechtungen sachlicher Art und auch Zeitreihen vorzuhalten und für den Direktzugriff anzubieten. In GENESIS-Online werden die komplexen Gebilde der Datenquader zur Vereinfachung als „Datensätze“ bezeichnet.

Die Funktionen

Die Entwicklung von GENESIS-Online folgte der Zielsetzung, es jedem Nutzer zu ermöglichen, das breite und tief gegliederte Datenangebot der amtlichen Statistik in komfortabler Weise über das Internet zu erschließen. Um dieses Ziel zu erreichen, stellt GENESIS-Online ausgewählte GENESIS-Funktionen über eine HTML-Oberfläche (siehe Schaubild 2) für den externen Nutzer zur Verfügung, die nachfolgend im Einzelnen beschrieben werden.

Schaubild 2



Recherche

GENESIS-Online bietet zahlreiche Recherchemöglichkeiten, um Daten und ihre beschreibenden Elemente, die Metadaten, zu finden.

Über die Begriffsrecherche können bis zu fünf Suchbegriffe durch „und“ (d. h. die Treffer enthalten jeden Begriff) bzw. „oder“ (d. h. die Treffer enthalten mindestens einen der Begriffe) miteinander verknüpft werden. Das Rechercheergebnis wird in objektspezifischen Trefferlisten (Statistiken, Merkmale, Tabellen, Zeitreihen, Daten) angezeigt. Die Suche nach einem einzelnen Begriff kann auch direkt über das Suchfeld im linken Menü vorgenommen werden.

Die Grundlage der Begriffsrecherche bildet ein Thesaurus, der auf die Datenbankinhalte abgestimmte Schlagworte (Begriffe) enthält. Berücksichtigt werden hierbei auch so genannte „Synonyme“, das heißt ähnliche Begriffe aus dem allgemeinen Sprachgebrauch, die bei der Suche automatisch auf den entsprechenden in der Datenbank vorhandenen „amtlichen“ Begriff verweisen und somit ebenfalls zu Treffern führen. Die Möglichkeit, synonyme Begriffe durch Beziehungen miteinander in Verbindung zu setzen, ist für die Effizienz der Recherche von besonderer Bedeutung. Durch die Aufnahme neuer Metadaten und die Auswertung der von den Nutzern eingegebenen Begriffe wird der Thesaurus sukzessive erweitert und verbessert. Die Liste aller im Thesaurus enthaltenen Begriffe wird dem Nutzer als Hilfe-

stellung über ein Buchsymbol hinter den Eingabefeldern für die Suchbegriffe zur Verfügung gestellt.

Neben der Begriffsrecherche werden noch eine Reihe verzeichnisorientierter Recherchemöglichkeiten angeboten:

Eine Recherche kann auch hierarchisch über die Sachgebiete bis zur Ebene der Statistiken erfolgen. Grundlage des hierarchischen Aufbaus der Sachgebiete und Statistiken ist das einheitliche Verzeichnis aller Statistiken der Statistischen Ämter des Bundes und der Länder (EVAS). Zu jeder Statistik werden weiterführende Informationen angeboten: Beschreibung der Statistik einschließlich ihrer Rechtsgrundlagen sowie Listen aller Merkmale, Datensätze und Tabellen, die zu der Statistik angeboten werden.

Eine weitere Möglichkeit zu recherchieren bietet der Merkmalskatalog, der alle in GENESIS-Online angebotenen Merkmale enthält. Die Merkmale können nach verschiedenen Kriterien angezeigt und sortiert werden und führen zu weiteren Informationen: Erläuterungen, Ausprägungen (bei klassifizierenden Merkmalen), sowie Statistiken, Datensätze und Tabellen, die das gewählte Merkmal nachweisen.

Schließlich gibt es auch noch Kataloge, die einen Überblick über alle angebotenen Tabellen bzw. Datensätze bieten. Sowohl Tabellen als auch Datensätze können über eine Identifikationsnummer, die in den ersten fünf Stellen den Code der zugrunde liegenden Statistik enthält, selektiert werden.

Tabellen

GENESIS-Online ermöglicht den Nutzern über flexible Tabellenstrukturen (Abruftabellen; siehe Schaubild 3) auf die Daten zuzugreifen, das heißt der Nutzer kann die Ausprägungen von vorgegebenen Merkmalen oder sogar die Merkmale für eine bestimmte Tabellenposition selbst auswählen. Variierbar sind zum Beispiel der gewünschte Zeitraum und die Auswahl von Positionen aus fachlichen und regionalen Gliederungen. Über eine Vorschau kann der individuell modifizierte Tabellenrahmen angezeigt werden. Der Wertabruf kann aber auch direkt gestartet werden, ohne dass der Nutzer individuelle Anpassungen vornimmt. Eine Ergebnistabelle (siehe Schaubild 4) wird dann nach den Vorgaben der Tabellenstruktur dynamisch aus einem oder mehreren gespeicherten Datensätzen erzeugt, sodass immer der aktuelle Stand der Datenbank wiedergegeben wird. Dies ist ein wesentlicher Vorteil im Vergleich zu Angeboten mit vordefinierten statischen Tabellen. Die Ergebnistabelle kann direkt in den Ausgabeformaten HTML, CSV und Excel® auf den eigenen PC übertragen werden (Download).

Größere Tabellen werden im Hintergrund (Batchbetrieb) erstellt und können später (i. d. R. nach wenigen Minuten) unter dem Menüpunkt „Aufträge/Ergebnisse“ abgeholt werden. Die im Batchbetrieb erzeugten Ergebnistabellen stehen dem Nutzer sieben Tage zur Abholung zur Verfügung. In dem Verzeichnis „Meine Tabellen“ können häufig benötigte und gegebenenfalls durch getroffene Auswahlen individuell angepasste Tabellenstrukturen vom Nutzer gespeichert werden.

Zeitreihen

Seit Juli 2004 bietet GENESIS-Online auch die Möglichkeit, die gespeicherten Datensätze (Quader) als Zeitreihen in einer standardisierten Struktur tabellarisch darzustellen.

Der zeitreihenorientierten Darstellung der statistischen Ergebnisse liegt ein komplexer Algorithmus zugrunde, der die Linearisierung eines Datensatzes (Quaders) bewirkt. Das Ergebnis der Linearisierung ist eine Ergebnistabelle mit standardisierter Struktur in Form von Zeitreihen, wobei die Zeitangaben grundsätzlich im Kopf der Tabelle ange-

Schaubild 3

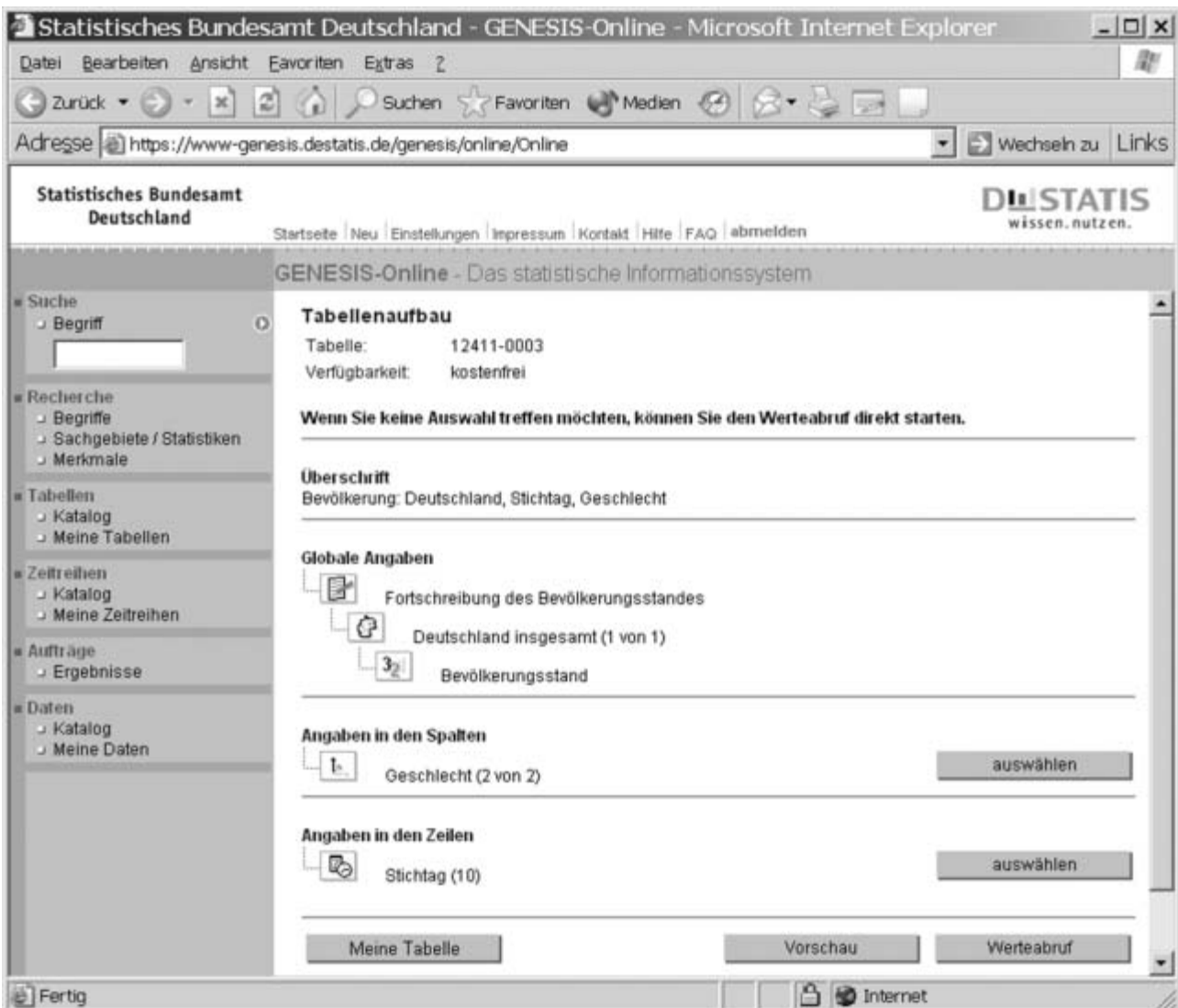
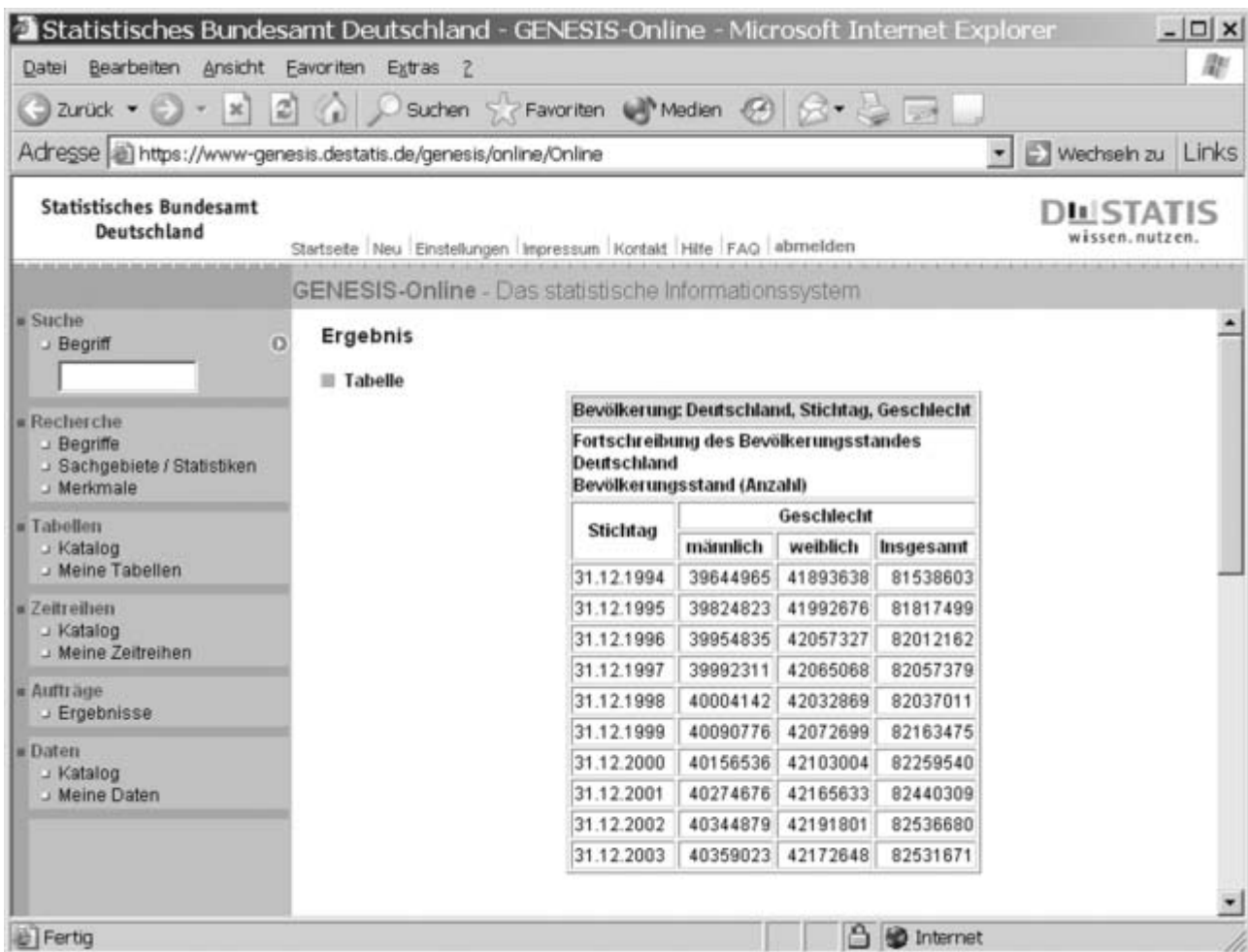


Schaubild 4



ordnet sind. Die Vorspalte der Tabelle enthält in hierarchischer Unterordnung Regionalmerkmal(e) und sachklassifizierende Merkmale. Den klassifizierenden Merkmalen sind auf der niedrigsten Stufe die Wertmerkmale in Nebenordnung untergeordnet.

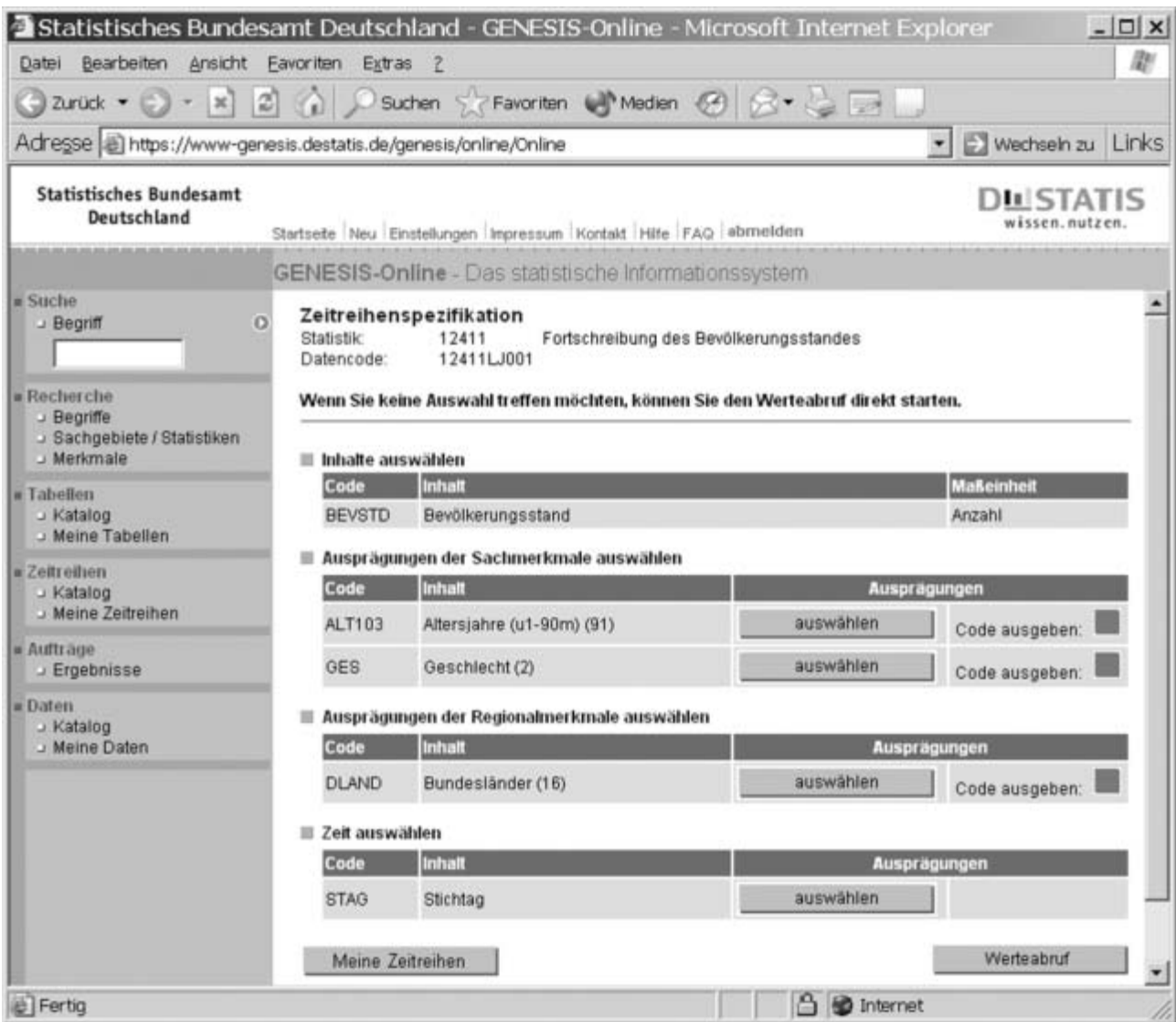
Nach der Auswahl eines gewünschten Datensatzes (Quaders) hat der Nutzer die Möglichkeit, eine Auswahl der Inhalte (Wertmerkmale), der Ausprägungen von Sach- und Regionalmerkmalen sowie der Zeit vorzunehmen (siehe Schaubild 5 auf S. 1106). Existieren zu einem Sachmerkmal hierarchisch übergeordnete Daten, so können die übergeordneten Hierarchiestufen, die in separaten Datensätzen (Quadern) gespeichert sind, hinzugefügt werden. Hierdurch wird die Möglichkeit eröffnet, die Informationen mehrerer, hierarchisch zusammenhängender Datensätze (Quader) in einer Ergebnistabelle zusammenzufassen. Bei Bedarf können auch die Fachschlüssel (Codes) von Merkmalsausprägungen ausgegeben werden, was zum Beispiel bei Klassifikationen (von Wirtschaftszweigen, Gütern) empfehlenswert ist. Erläuterungen zum Datensatz werden als Fußnote grundsätzlich automatisch mit ausgegeben.

Da in GENESIS die tabellarische Darstellung nach rechts (maximal 120 Spalten) und nach unten (maximal 99999 Zeilen) begrenzt ist und zudem in der Vorspalte nur eine

begrenzte Anzahl von Merkmalen dargestellt werden kann, ist bei langen Reihen und umfangreichen Gliederungen unter Umständen eine zeitliche oder auch eine inhaltliche Segmentierung notwendig. Überschreiten die gespeicherten Zeitangaben die Anzahl der verfügbaren Spalten, so werden die maximal darstellbaren Zeitpunkte vom System automatisch ermittelt und dem Nutzer angezeigt. Wird die Anzahl der verfügbaren Zeilen überschritten, kann der Nutzer individuell eine entsprechende Auswahl der Wertmerkmale und/oder Auswahl von Ausprägungen klassifizierender Merkmale vornehmen. Ist die Kapazität der Vorspalte hinsichtlich der Anzahl darstellbarer Merkmale erschöpft, so kann zum Beispiel durch die Auswahl von genau einer Ausprägung eines klassifizierenden Merkmals bzw. im Fall mehrerer Wertmerkmale durch die Auswahl genau eines Wertmerkmals die Vorspalte inhaltlich reduziert werden, sodass die Erstellung der Ergebnistabelle möglich wird. Durch entsprechende Systemmeldungen wird der Nutzer nach dem Start des Werteabrufs auf eventuell erreichte Grenzen hingewiesen, um dann entsprechende Anpassungen vornehmen zu können.

Nach dem Werteabruf kann die Ergebnistabelle dann direkt in den Ausgabeformaten HTML, CSV und Excel® auf den eigenen PC übertragen werden (Download). Größere Tabellen werden auch hier im Hintergrund (Batchbetrieb) erstellt

Schaubild 5



und können zu einem späteren Zeitpunkt unter „Aufträge/Ergebnisse“ abgeholt werden. Die durch den Nutzer individuell spezifizierten Zeitreihen können zum wiederholten Abruf in einem eigenen Verzeichnis („Meine Zeitreihen“) gespeichert werden.

Daten

GENESIS-Online bietet die Möglichkeit, die Datensätze (Quader) in einem linearisierten Format zu exportieren. Auf Wunsch können auch die zugehörigen Metadaten exportiert werden. Dies ist besonders für die Nutzer von Interesse, die die Daten in eigenen Datenbanken weiterverarbeiten möchten. Durch getroffene Auswahlen individuell spezifizierte Datenexporte können vom Nutzer auch hier in einem eigenen Verzeichnis („Meine Daten“) gespeichert werden.

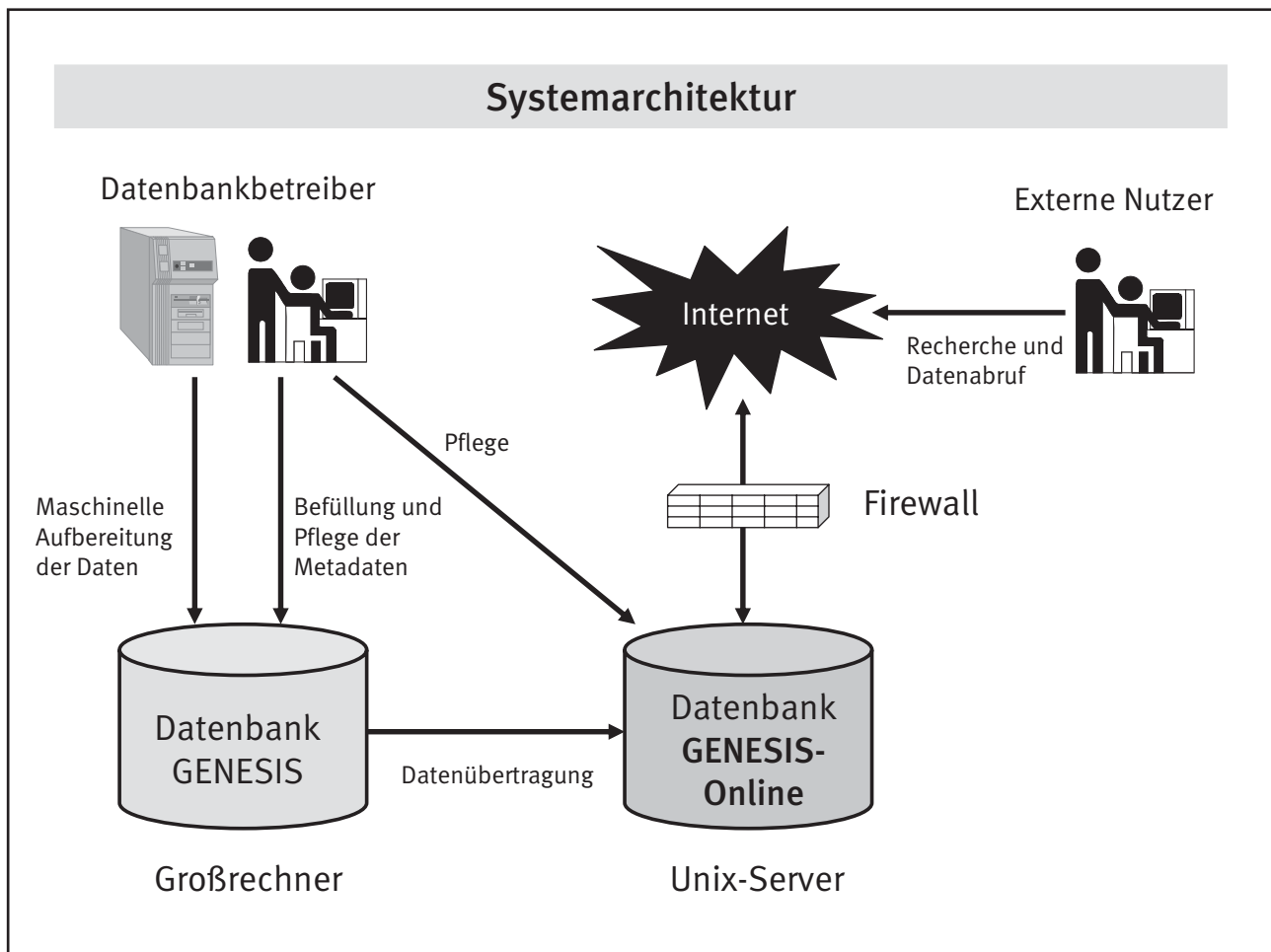
Systemarchitektur

GENESIS-Online wird über das interne Datenbanksystem GENESIS als zentraler Serverkomponente befüllt und aktu-

alisiert. Als Benutzerschnittstellen besitzt GENESIS neben einer klassischen zeichenorientierten Oberfläche eine graphische Oberfläche (Java-Client) für die interne Nutzung des Systems sowie eine HTML-Oberfläche (HTML-Client) für die Nutzung durch externe Kunden über das Internet. Die Befüllung von GENESIS mit Daten erfolgt maschinell. Für die Pflege der Datenbanken GENESIS und GENESIS-Online steht die zeichenorientierte oder auch die graphische Oberfläche mit einer umfangreichen Funktionalität zur Verfügung. Die in Schaubild 6 schematisch dargestellte Systemarchitektur bietet die Möglichkeit, die Daten effizient und mit bestmöglicher Aktualität im Internet bereitzustellen.

Mit GENESIS-Online wurde eine Oberfläche entwickelt, die auf der Grundlage einer XML-Schnittstelle die Datenbestände und Funktionalität von GENESIS erschließt. GENESIS-Online enthält keine statischen Seiten, sondern generiert direkt entsprechend der Nutzeranforderung ein aktuelles Abbild der gespeicherten Information. Die hierzu eingesetzte Technologie (Java, XML, Middleware) ist zukunftsicher und bietet ein großes Potenzial für Erweiterungen.

Schaubild 6



Datenangebot

Das Datenangebot in GENESIS-Online ist unter <http://www.destatis.de/genesis> abrufbar und umfasst Daten der amtlichen Statistik aus folgenden Sachgebieten:

- Gebiet, Bevölkerung, Erwerbstätigkeit, Wahlen
- Bildung, Sozialleistungen, Gesundheit, Recht
- Wohnen, Umwelt
- Wirtschaftsbereiche
- Außenhandel, Unternehmen, Handwerk
- Preise, Verdienste, Einkommen und Verbrauch
- Öffentliche Finanzen
- Gesamtrechnungen.

Der Menüpunkt „Neu“ informiert über die laufenden Datenaktualisierungen und -ergänzungen in GENESIS-Online. Das Datenangebot umfasst aktuell etwa 30 Mill. Werte und wird weiterhin sukzessive unter Berücksichtigung der Kunden-

wünsche ausgebaut. Für den outputorientierten Ausbau der Inhalte werden insbesondere die von den Nutzern bei der Begriffsrecherche angefragten Begriffe laufend ausgewertet und das Datenangebot damit verbessert.

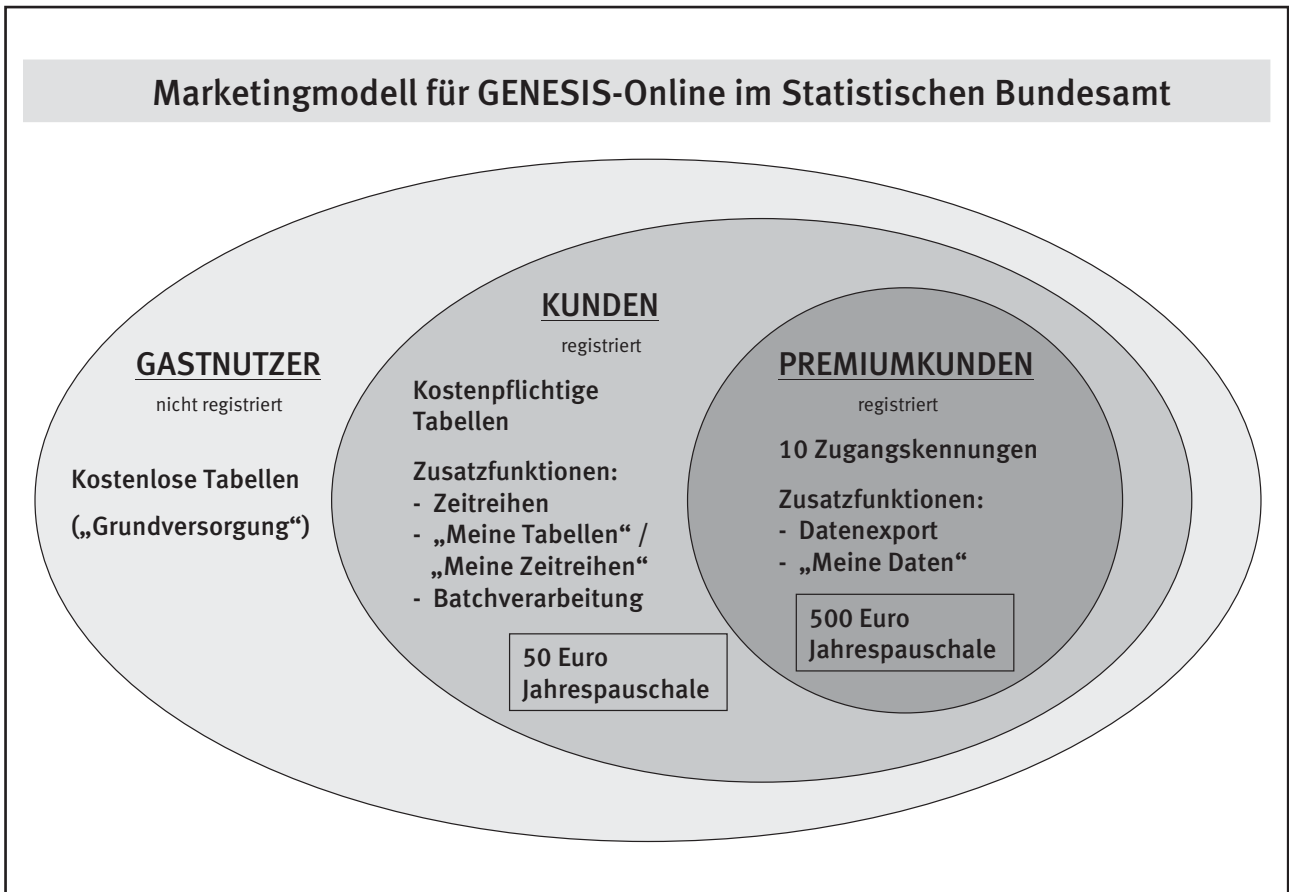
Zugangswege und Nutzung

Nach der Freischaltung von GENESIS-Online im Juli 2002 folgte eine zweijährige Aufbauphase, in der zahlreiche Nutzer das System kostenlos testen konnten. In dieser Zeit wurden wertvolle Hinweise und Wünsche der Nutzer aufgenommen und umgesetzt. GENESIS-Online wurde 2004 um einen kostenpflichtigen Bereich mit tief gegliederten Tabellen und zusätzlichen Funktionen erweitert und hat damit den Internetdienst „Zeitreihenservice“ abgelöst.

Die unterschiedlichen Anforderungen der verschiedenen Nutzergruppen führten zu der Entwicklung eines Marketingmodells, das den Nutzern verschiedene Zugangswege als Gäste, Kunden und Premiumkunden eröffnet (siehe Schaubild 7 auf S. 1108).

Im Rahmen der informationellen Grundversorgung hat jeder als Gastnutzer freien Zugang zum kostenlosen Tabellenangebot und zur umfassenden Datenrecherche.

Schaubild 7



Registrierten Kunden stehen gegen eine jährliche Gebühr von 50 Euro das vollständige Datenangebot und zusätzliche Funktionen zur Verfügung: Zugang zu den kostenpflichtigen Tabellen (gekennzeichnet mit dem Euro-Symbol), Erstellung von Zeitreihen sowie individualisierte Funktionen wie „Meine Tabellen“, „Meine Zeitreihen“ und die Hintergrundverarbeitung (Batchbetrieb) für den Abruf großer Tabellen.

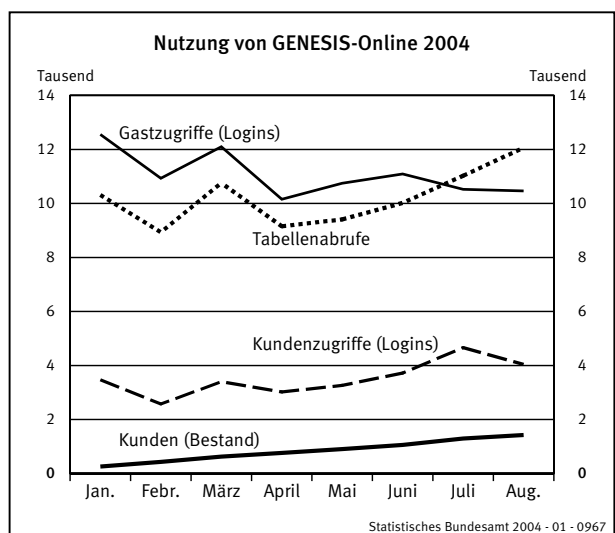
Geschäftskunden – zum Beispiel Verbände oder große Unternehmen (so genannte Premiumkunden) – erhalten für 500 Euro pro Jahr neben den Nutzungsmöglichkeiten für Kunden zusätzliche Leistungen: Sie können mit zehn Zugangskennungen die Datenbestände über die Exportfunktion aus GENESIS-Online beziehen und in ihren eigenen Datenbanken weiterverarbeiten.

Zur kostenpflichtigen Registrierung steht auf der Anmelde-seite von GENESIS-Online ein Registrierungsformular zur Verfügung. Nach der Registrierung erhält der Nutzer vom Statistischen Bundesamt eine E-Mail mit einer Zugangskennung für GENESIS-Online, bestehend aus einem Nutzernamen und einem Passwort.

Von Januar bis August 2004 konnten 1 400 Kunden und 34 Premiumkunden gewonnen werden. Im gleichen Zeitraum wurden 88 500 Gastanmeldungen gezählt. Seit Januar 2004 finden mit steigender Tendenz pro Monat durchschnittlich 10 000 Tabellenabrufe statt. Schaubild 8 zeigt die Entwicklung des über die Zeit kumulierten Kundenbestan-

des, die Anzahl der monatlichen Gast- und Kundenzugriffe (Logins) sowie die Anzahl der Tabellenabrufe in den einzelnen Monaten.

Schaubild 8



Bei fachlichen oder administrativen Fragen finden Nutzer unter dem Menüpunkt „Kontakt“ zahlreiche Ansprechpartner. Eine Liste mit der Beantwortung der am häufigsten ge-

stellten Fragen (siehe Menüpunkt „FAQ“ = Frequently asked questions) rundet den Service ab.

Fazit

Mit GENESIS-Online wurde eine benutzerfreundliche und innovative Software entwickelt, die es den Statistischen Ämtern des Bundes und der Länder ermöglicht, ihre Daten über eine einheitliche Benutzeroberfläche über das Internet bereitzustellen. Neben dem Einsatz im Statistischen Bundesamt wird die Software von GENESIS-Online bisher auch in dem gemeinsamen Statistik-Portal der Statistischen Ämter (GENESIS-Online regional) und im Bayerischen Landesamt für Statistik und Datenverarbeitung (GENESIS-Online-Bayern) eingesetzt. [u](#)

Auszug aus Wirtschaft und Statistik

© Statistisches Bundesamt, Wiesbaden 2004

Für nichtgewerbliche Zwecke sind Vervielfältigung und unentgeltliche Verbreitung, auch auszugsweise, mit Quellenangabe gestattet. Die Verbreitung, auch auszugsweise, über elektronische Systeme/Datenträger bedarf der vorherigen Zustimmung. Alle übrigen Rechte bleiben vorbehalten.

Herausgeber: Statistisches Bundesamt, Wiesbaden

Schriftleitung: Johann Hahlen
Präsident des Statistischen Bundesamtes
Verantwortlich für den Inhalt:
Brigitte Reimann,
65180 Wiesbaden

- Telefon: +49 (0) 6 11/75 20 86
- E-Mail: wirtschaft-und-statistik@destatis.de

Vertriebspartner: SFG Servicecenter Fachverlage
Part of the Elsevier Group
Postfach 43 43
72774 Reutlingen
Telefon: +49 (0) 70 71/93 53 50
Telefax: +49 (0) 70 71/93 53 35
E-Mail: destatis@s-f-g.com

Erscheinungsfolge: monatlich



Allgemeine Informationen über das Statistische Bundesamt und sein Datenangebot erhalten Sie:

- im Internet: www.destatis.de

oder bei unserem Informationsservice
65180 Wiesbaden

- Telefon: +49 (0) 6 11/75 24 05
- Telefax: +49 (0) 6 11/75 33 30
- E-Mail: info@destatis.de