

GENESIS-Online (Bayern) – Das statistische Informationssystem im Internet

Dipl.-Geogr. Michael Carle

Mit der Einführung der Online-Datenbank GENESIS-Online zum Jahresbeginn 2004 wurde ein wichtiger Beitrag zum wachsenden Datenbedarf im Rahmen der heutigen Informationsgesellschaft geleistet sowie die zunehmend bedeutende Rolle des Internets als modernes Kommunikationsmedium erkannt und genutzt. GENESIS ist ein von den statistischen Ämtern des Bundes und der Länder gemeinsam entwickeltes statistisches Informationssystem (Gemeinsames Neues Statistisches InformationSystem), das durch die Entwicklung von GENESIS-Online zu einer internetfähigen Client/Server-Anwendung ausgebaut wurde. Mit dieser Auskunftsdatenbank bietet das Bayerische Landesamt für Statistik und Datenverarbeitung allen interessierten Nutzern die Möglichkeit, sich das breit gefächerte Datenangebot der GENESIS-Datenbank des Landesamts jederzeit im Internet zu erschließen. Dabei werden die Daten über GENESIS-Online im Rahmen der informationellen Grundversorgung kostenlos zur Verfügung gestellt. Im Folgenden werden Inhalt, Aufbau und Funktionalitäten der Datenbank näher erläutert, sowie einige Nutzungszahlen dargestellt.

Datenangebot

Das Datenangebot in GENESIS-Online umfasst zurzeit ca. 370 Tabellen zu etwa 60 Statistiken aus folgenden Bereichen:

- Gebiet, Bevölkerung, Erwerbstätigkeit, Wahlen
- Bildung, Sozialleistungen, Gesundheit
- Wohnen
- Wirtschaftsbereiche
- Außenhandel, Unternehmen
- Öffentliche Finanzen

Daten-
grundlage

Als Datengrundlage für GENESIS-Online dient die neue Produktionsdatenbank GENESIS. Dort sind derzeit rund 115 Mill. Werte gespeichert die einerseits in einem komplexen Übernahmeprozess aus der bisherigen IMS Datenbank übernommen wurden und zum anderen aus Datenlieferungen auf Basis des Gemeinschaftsprojekts GENESIS¹ der statistischen Ämter des Bundes und der Länder stammen.

Neben der laufenden Aktualisierung dieses Angebots besteht natürlich die Bestrebung der inhaltlichen und funktionalen Erweiterung im Rahmen der Anpassung an den Bedarf der Nutzer der amtlichen Statistik.

Daten-
bereit-
stellung

Die Bereitstellung der Daten in GENESIS-Online (Bayern) erfolgt mittels vordefinierter Tabellen, so genannter Standardtabellen. Variable Elemente ermöglichen es, alternative Merkmale zu wählen oder auch gezielt Ausprägungen auszuwählen und so die Tabelle flexibel an den jeweiligen Bedarf anzupassen.

Tabellenkonzept

Im Rahmen der Erstellung des Tabellenangebots wurde zunächst ein Konzept erarbeitet, welches folgende Kriterien berücksichtigen sollte:

- Orientierung am Bedarf des Nutzers
Die im Auskunftsdienst eingehenden Anfragen zeigen, dass es zwei grundsätzliche Sichtweisen bei regionalisiertem Datenbedarf gibt:
 - „viele“ Daten einer Regionaleinheit;
 - alle Regionaleinheiten in einem Abruf, dann meist nur für ein oder zwei Berichtsjahre, keine Zeitreihen.
- Die Tabellen sollen übersichtlich sein.
- Die Tabellen sollten vorwiegend so gestaltet sein, dass die Datenmenge online abgerufen werden kann.
- Die abrufbaren Datenmengen sollten die Rechnerlast berücksichtigen; ein Abruf der Bevölkerung nach Geschlecht, Nationalität, Altersjahren und Familienstand in den Regierungsbezirken für ein Jahr hat bereits 36 360 Werte. Selbst „kleine“ Gemeindetabellen haben bereits 6 000 bis 10 000 Werte.

Um diese Kriterien optimal zu berücksichtigen, differenziert das Konzept des Tabellenangebots zwischen folgenden Tabellentypen:

- Eckzahlentabellen
Sie bieten für Eckzahlen – Wertmerkmale ohne weitere sachliche Gliederung wie z.B. Bevölkerung insgesamt – sowohl ei-

Tabellen-
typen

¹ Siehe Bayern in Zahlen 1/2002: „GENESIS ante portas“

ne tiefe regionale Gliederung in der Vorspalte, als auch eine begrenzte Zeitreihe im Tabellenkopf an.

- **Regionaltabellen**
Sie bieten Daten in regionaler und sachlicher Gliederung für alle Regionaleinheiten, jedoch nur *eine* Zeitausprägung an. Dabei ist die Sachgliederung meist dem Tabellenkopf zugeordnet, die Regionalgliederung der Vorspalte, z.B. Bevölkerung nach Geschlecht auf Gemeindeebene. Eine weitere Sach- oder Zeitgliederung in der Vorspalte bleibt wegen der damit verbundenen großen Abrufmengen die Ausnahme.
- **Zeitreihentabellen**
Sie bieten die Daten für mehrere Jahre an, jedoch nur für *eine* Regionaleinheit je Abruf.
- **Strukturtabellen**
Sie bieten schwerpunktmäßig Sachgliederungen, jedoch nicht nur im Tabellenkopf wie bei den vorangegangenen Tabellentypen, sondern auch in den übrigen Strukturelementen wie Vorspalte und Zwischenüberschrift. Die Regional- und Zeitauswahl ist grundsätzlich auf jeweils *eine* Ausprägung beschränkt, da diese Tabellen ansonsten schnell recht groß werden.

Tabellencode Zu jeder Tabelle werden der Tabellencode sowie die Überschrift angezeigt. Dabei besteht der Tabellencode aus dem fünfstelligen EVAS-Schlüssel² der jeweiligen Statistik sowie einem Buchstabenkürzel, das den Tabellentyp bestimmt:

- r = Regionaltabelle
- z = Zeitreihentabelle
- s = Strukturtable
- kein Kürzel = Eckzahlentabelle

Aufbau der Datenbank

Das nach Statistiken geordnete Datenangebot folgt in der Struktur des Statistikverzeichnisses den Codes des einheitlichen Verzeichnisses aller Statistiken (EVAS).

Die gespeicherten Werte zu den einzelnen Statistiken sind zeitlich, räumlich sowie sachlich durch die so genannten Metadaten näher definiert.

Metadaten Die wichtigsten Metadaten sind

- Wertmerkmale,
- zeitlich, räumlich und sachlich klassifizierende Merkmale sowie
- zeitidentifizierende Merkmale.

Wertmerkmale beschreiben den abrufbaren Inhalt, der durch eine zu definierende Zeiteinheit sowie eine räumliche Zuordnung näher eingegrenzt wird. Eine weitere Konkretisierung des Datenbankinhalts wird durch sachliche Gliederungen hergestellt.

Demnach besitzen die zeitlich, räumlich und sachlich klassifizierenden sowie die zeitidentifizierenden Merkmale mehrere Merkmalsausprägungen, die den Inhalt tiefer strukturieren.

Beispiel:

Wird aus der Datenbank die Information benötigt, wie viele weibliche Personen eines bestimmten Alters und Nationalität innerhalb eines bestimmten bayerischen Regierungsbezirks zu einem bestimmten Stichtag leben, so wird die Abfrage durch folgende Datenbankobjekte definiert:

- Statistik; hier: 12 411 Fortschreibung des Bevölkerungsstandes
- Merkmale:
 - Wertmerkmal; hier: „Bevölkerung“
 - räumlich klassifizierendes Merkmal; hier: Regierungsbezirke Bayerns
 - sachlich klassifizierende Merkmale; hier: Alter, Nationalität, Geschlecht
 - zeitidentifizierendes Merkmal; hier: Stichtag (z.B. 31.12.2004)

Dabei besitzen die klassifizierenden Merkmale sowie das zeitidentifizierende Merkmal folgende Merkmalsausprägungen:

Regierungsbezirke Bayerns – Oberbayern, Niederbayern, ..., Schwaben

Stichtag – 31.12.1985, 31.12.1986, ..., 31.12.2004

Alter – 1 Jahr, 2 Jahre, ..., 100 oder mehr Jahre

Nationalität – Ausländer, Deutsche

Zugang und Funktionen

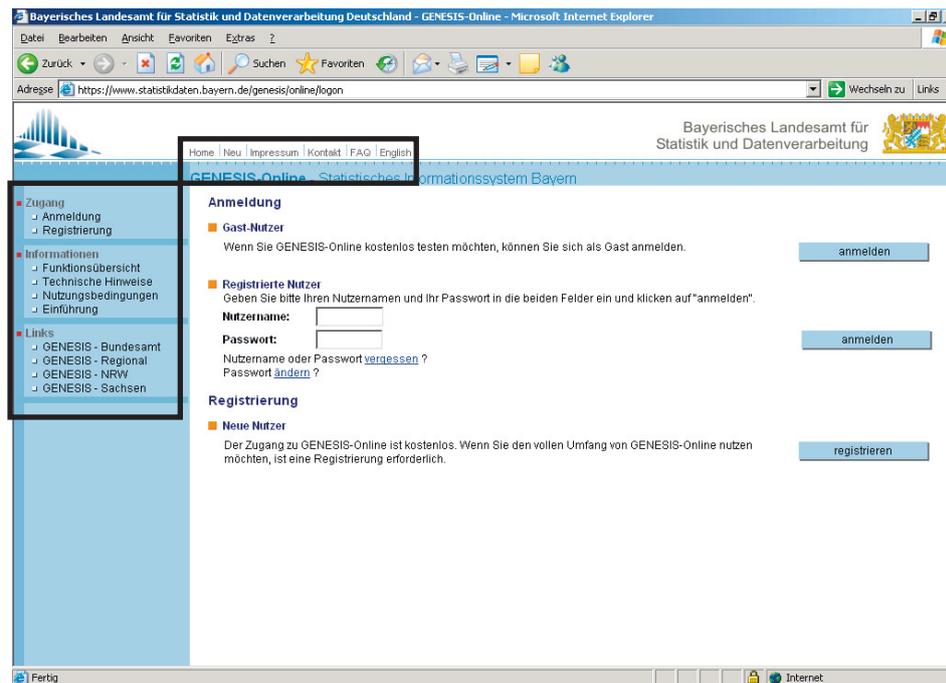
GENESIS-Online (Bayern) ist über einen Link auf der Homepage des Bayerischen Landesamts für Statistik und Datenverarbeitung (<http://www.statistik.bayern.de>) oder direkt über <http://www.statistikdaten.bayern.de> erreichbar. Die Datenbank kann ebenso als Gast sowie als registrierter Kunde genutzt werden. Dabei unterscheidet sich die Nutzung als Gast gegenüber jener als registrierter Kunde durch eine eingeschränkte Funktionalität sowie eine geringere abrufbare Datenmenge. Eine Registrierung ist nicht mit Kosten verbunden, da GENESIS, wie bereits erwähnt, im Rahmen der informationellen Grundversorgung generell kostenfrei angeboten wird;

² EVAS = Einheitliches Verzeichnis aller Statistiken der statistischen Ämter des Bundes und der Länder.

sie ist jedoch erforderlich, da nur registrierte Nutzer größere Datenmengen abrufen sowie eine in der Tabellenstruktur vorgenommene individuelle Auswahl abspeichern können. Diese steht dem Nutzer dann bei Bedarf immer wieder zur Verfügung.

Im Folgenden werden die einzelnen Funktionen der Datenbank genauer beschrieben:

Allgemeine Informationsangebote



Information für den Nutzer

Bereits auf der Anmeldeseite stehen dem Nutzer wichtige Informationen zur Nutzung und zu aktuellen Änderungen bzw. Erweiterungen der Datenbank zur Verfügung (siehe markierte Bereiche).

Unter dem Punkt „Informationen“ am linken Bildschirmrand können Erläuterungen zu Funktionen, zu den notwendigen technischen Voraussetzungen zur optimalen Systemnutzung und allgemeine Nutzungsbedingungen angezeigt werden.

Hinter der Rubrik „Einführung“ verbirgt sich ein graphischer Wegweiser, der den Einstieg und den Datenabruf Schritt für Schritt erläutert.

Über den Punkt „Links“ gelangt man zu den GENESIS-Online Datenbanken anderer Bundesländer sowie des Statistischen Bundesamts.

Mit GENESIS-Online Regional wird auf die GENESIS Datenbank im Informationsangebot des gemeinsamen Statistik-Portals der statistischen Ämter des Bundes und der Länder verlinkt. Hier fin-

det man Daten zu allen Bundesländern in tiefer regionaler Gliederung bis zur Kreisebene. Datengrundlage ist der regionalstatistische Datenkatalog des Bundes und der Länder.

Im oberen Bildschirmbereich können über den Menüpunkt „Neu“ Informationen über Datenaktualisierungen sowie Erweiterungen des Tabellenangebots oder die Aufnahme neuer Statistiken abgerufen werden. Neben dem Impressum und den Angaben zur

Kontaktaufnahme bei Fragen oder Anregungen stehen unter dem Menüpunkt „FAQ“ die häufig gestellten Fragen (FAQ – Frequently Asked Questions) und deren Beantwortung als erste Anlaufstelle für Fragen zur Verfügung.

Auch besteht die Möglichkeit, sich die Menüführung in englischer Sprache anzeigen zu lassen; allerdings sind die Begrifflichkeiten der Statistik hierbei nicht mit einbezogen.

Recherchemöglichkeiten

Das System bietet verschiedene Möglichkeiten, die Inhalte nach der gewünschten Information zu durchsuchen. Die Recherchefunktionen stehen dem Gastnutzer ebenso zur Verfügung, wie dem registrierten Nutzer.

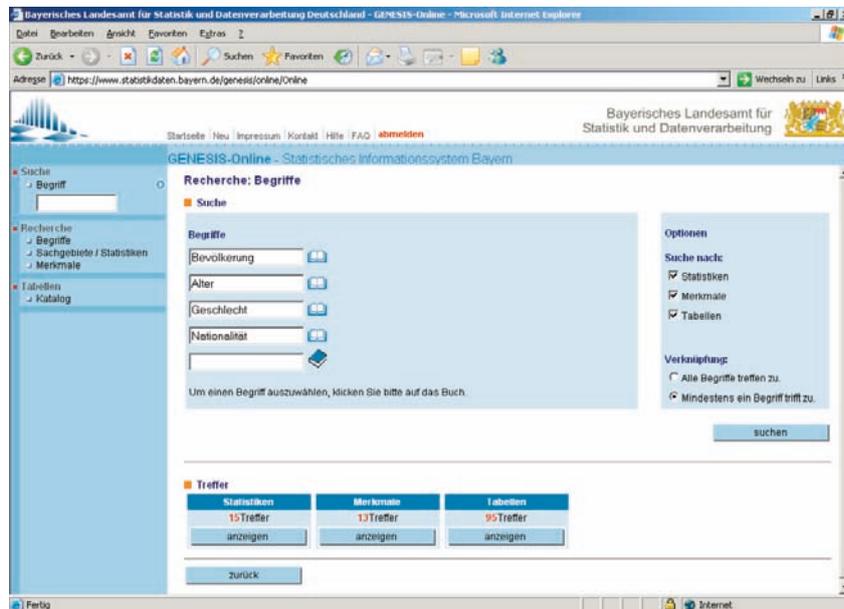
- die Thematische Recherche
- Grundlage dieser Recherche sind die hierarchisch nach der EVAS Nummer gegliederten „Sachgebiete“ und Statistiken.

Thematische Recherche

Ausgehend von den „Sachgebieten“ führt sie über die einzelnen Statistiken zu den jeweils verfügbaren Tabellen. Auf der Ebene der Statistiken werden Informationen zur Statistik und zu den in der Statistik gebrauchten Merkmalen angeboten.

Wortstamm ergänzen und das System veranlassen, alle diesen Bestandteil nutzenden Begriffe zu suchen, sind nicht möglich.

- die Suche über ein oder mehrere Stichworte



Begriffsrecherche

In die „Begriffsrecherche“ können bis zu fünf Begriffe mit einbezogen werden. Dabei hat man die Wahl entweder nach allen oder nach nur einem der Begriffe zu suchen („alle Begriffe treffen zu“; „mindestens einer der Begriffe trifft zu“). Bei der Eingabe der Begriffe kann auf eine Stichwortliste zurückgegriffen werden, die durch „aufklappen“ des blauen Buchsymbols hinter jeder Eingabezeile zugänglich ist. Weiter wird die Suche durch Angabe des/der zu suchenden Objekte näher definiert. Man kann nach Statistiken, Merkmalen sowie nach Tabellen zu einer Begriffseingabe suchen. Über Links im unteren Bildschirmbereich („Treffer“) wird dann der Zugriff zu den gefundenen Objekten hergestellt.

Die Recherche greift auf eine im System generierte Begriffsliste zu, die durch die Eingabe von Synonymen und Assoziationen, wie z.B. Import – Einfuhr, erweitert werden kann. Außerdem können die von Nutzern eingegebenen Begriffe, zu welchen kein Ergebnis gefunden wurde gelistet werden, um diese mit den statistikinternen Begrifflichkeiten zu verknüpfen. D.h. das System wächst mit jeder Suche um Begriffe, die anhand der Begriffslistung eingepflegt werden.

Dabei muss immer der vollständige Begriff bei der Suche angegeben werden, so genannte „Platzhalter“, die einen bestimmten

Wird nur nach einem bestimmten Begriff gesucht, so kann dies direkt vom Startbildschirm durch Eingabe in das Feld „Suche – Begriff“ geschehen.

- die Suche auf Basis der Merkmale

Als weitere Suchoption wird die Suche über den in der Datenbank vorgehaltenen Merkmalskatalog angeboten. Hier kann auf Basis der Anfangsbuchstaben die Liste der zugehörigen Merkmale aufgeschlagen werden, um das Gewünschte zu finden. Zur Eingrenzung der Suche kann auch der zu suchende Merkmalstyp (alle, nur Wertmerkmale, nur klassifizierende Merkmale) zuvor eingestellt werden.

Merkmalskatalog

Innerhalb des angezeigten Merkmalskatalogs sind Informationen in Form von Erläuterungen zum jeweiligen Merkmal abrufbar, ebenso wie Informationen zum Merkmalstyp oder dessen Ausprägungszahl. Auch kann direkt in das Statistikverzeichnis und Tabellenverzeichnis gewechselt werden, wobei dann genau die Statistiken oder Tabellen angezeigt werden, die das Merkmal verwenden.

Tabellenabrufe

Zentrales Element für den Zugriff auf die im System gespeicherten Daten sind die vordefinierten Tabellen.

Datenabruf

Neben den vielseitigen Recherchemöglichkeiten, die jeweils auf die Tabellen hinführen, kann über die Funktion „Katalog – Tabellen“ die Liste aller im System verfügbaren Tabellen angezeigt werden.

Tabellen-
beschreibung

Neben dem Tabellencode (siehe Tabellenkonzept) steht zu jeder Tabelle ein den Inhalt beschreibender Text zur Verfügung, der auch über die bestehenden Auswahlmöglichkeiten innerhalb der Tabelle Auskunft gibt:

Code	Inhalt
12411-007s	Bevölkerung: Kreis, Geschlecht, Altersjahre (75)/Altersjahre (88), Stichtag

Die Tabelle, die in diesem Beispiel dargestellt wird, hat die Bevölkerung zum Thema. Dieses wird immer als erstes genannt, danach folgt die nähere Inhaltsdefinition. Hier können Daten zu den bayerischen Kreisen sowie zu allen höheren administrativen Gebietseinheiten (Regierungsbezirke, Land) abgerufen werden. D.h. im Text wird immer die niedrigste zur Verfügung stehende Regionaleinheit, zu der Informationen abgerufen werden können angegeben. Dabei wird die Bevölkerung nach Geschlecht und Altersjahren zu den gespeicherten Stichtagen dargestellt. Hier kann zwischen zwei Altersgliederungen gewählt werden, wobei die Anzahl der Ausprägungen der unterschiedlichen Altersgliederungen in Klammern dahinter angezeigt wird. Der Schrägstrich zwischen den beiden Altersmerkmalen weist darauf hin, dass hier eine Auswahl möglich ist. In jeder Tabelle ist die regionale Gliederung sowie die Zeit flexibel wählbar.

- Strukturelemente

Wird eine Tabelle abgerufen, so gelangt man zunächst in das Fenster „Tabellenaufbau“, der die Tabelle in einzelne Strukturelemente unterteilt.

Es wird unterschieden zwischen den „Globalen Angaben“, den Angaben im Tabellenkopf („Angaben in den Spalten“) und den Angaben in der Vorspalte („Angaben in den Zeilen“). Diese sind im Schaubild mit Pfeilen markiert.

Die „Globalen Angaben“ sind dabei das Element, über das die abzurufende Datenmenge gesteuert werden kann; zu den hier platzierten klassifizierenden Merkmalen kann immer nur eine Merkmalsausprägung gewählt werden. Der Nutzer wird bei der Wahl eines anderen als dem voreingestellten Merkmal darauf hingewiesen, dass genau eine Ausprägung auszuwählen ist.

- Merkmals- und Ausprägungsauswahl

Im Rahmen der Auswahlmöglichkeiten gilt es zu unterscheiden, ob es sich um die Möglichkeit handelt, zwischen unterschiedlichen Merkmalen zu wählen, oder ob es um die Auswahl einzelner Ausprägungen innerhalb eines klassifizierenden Merkmals geht.

Auswahl-
möglichkeiten

Sofern an einer Tabellenposition mehrere Merkmale flexibel wählbar sind, steht eine Auswahlbox zur Verfügung, die durch „aufklappen“ die dort wählbaren Optionen anzeigt. Zu beachten ist, dass die in der „Box“ getätigte Auswahl erst durch Aktivierung der Schaltfläche „übernehmen“ (links neben der entsprechenden Tabellenposition) in den Abruf integriert wird.

Merkmals-
auswahl

Tabellen-
aufbau

The screenshot shows the 'Tabellenaufbau' (Table Structure) interface for the table 12411-007s. The interface is divided into three main sections, each with a red arrow pointing to it from the left sidebar:

- Globale Angaben (Global Information):** This section includes 'Fortschreibung des Bevölkerungsstandes' (Population Projection), 'Bevölkerung' (Population) with a dropdown for 'Bundesländer (1 von 1)' (Federal States) and a button 'auswählen' (select), and 'Stichtag (1)' (Reference Date) with a button 'auswählen'.
- Angaben in den Spalten (Information in Columns):** This section includes 'Geschlecht (2 von 2)' (Sex) with a button 'übernehmen' (take over).
- Angaben in den Zeilen (Information in Rows):** This section includes 'Altersjahre (unter 1 bis 95, 90, 95 oder älter) (88 von 88)' (Age Groups) with a dropdown menu and a button 'übernehmen'.

At the bottom of the interface, there are buttons for 'Vorschau' (Preview), 'Werteabruf' (Value Retrieval), and 'zurück' (Back).

Ausprägungs-
auswahl

Stehen bezüglich eines klassifizierenden Merkmals Ausprägungen zur Auswahl, so wird dies durch die Schaltfläche „auswählen“ angezeigt. Durch Aktivierung der Schaltfläche wird eine Liste der Ausprägungen angezeigt, aus der der Nutzer seine Auswahl treffen kann. Die Übernahme erfolgt auch hier durch die Schaltfläche „übernehmen“ am Ende der Seite.

Ist die Anzahl der Ausprägungen sehr groß, kann die Anzeige durch Eingabe eines Begriffs im Feld „Auswahl“ eingeschränkt werden. Hierbei besteht die Möglichkeit mit einem * als Platzhalter zu arbeiten. Wichtig ist innerhalb der Suchfunktion das richtige Suchkriterium einzustellen.

Die Möglichkeiten der Individualisierung des Abrufs durch Auswahl der Merkmale und Merkmalsausprägungen stehen dem Gast sowie dem registrierten Benutzer in gleicher Weise zur Verfügung.

- Meine Tabellen

Eigener Tabellenbereich
registrierter Nutzer

Diese Funktion steht nur registrierten Benutzern zur Verfügung, da es hier möglich ist, auf einmal getätigte Tabelleneinstellungen zum erneuten Abruf in einer späteren Sitzung wieder zurückzugreifen. Damit dies funktionieren kann, muss das System einen Nutzer bzw. eine Kennung haben, dem bzw. der es die erstellte Tabelle zuweisen kann.

Z.B. können sich hier Gemeinden Tabellen so abspeichern, dass die eigene Gemeinde beim erneuten Abruf nicht noch einmal extra gesucht werden muss. Die Tabellenstruktur wird in einem eigenen

Tabellenbereich abgelegt. Über den Menüpunkt „Meine Tabellen“ kann der jeweilige Tabellenbereich wieder aufgerufen werden.

- Abruf-, Anzeige-, Druck- und Exportmöglichkeiten
- Nach erfolgter Einstellung der gewünschten Merkmale bzw. Merkmalsausprägungen innerhalb des Tabellenaufbaus wird mit der Schaltfläche „Werteabruf“ die Datenabfrage gestartet.

Abruf- und
Verarbeitungsmöglichkeiten

Der Abruf wird in Form der so genannten „Ergebnistabelle“ angezeigt und kann weiter bearbeitet werden. Zum einen kann die Tabelle in kompakter, weiterverarbeitbarer Form, d.h. ohne mögliche „leere“ Zeilen und Spalten dargestellt werden oder aber, z.B. für die Druckausgabe vollständig in einem eigenen Fenster.

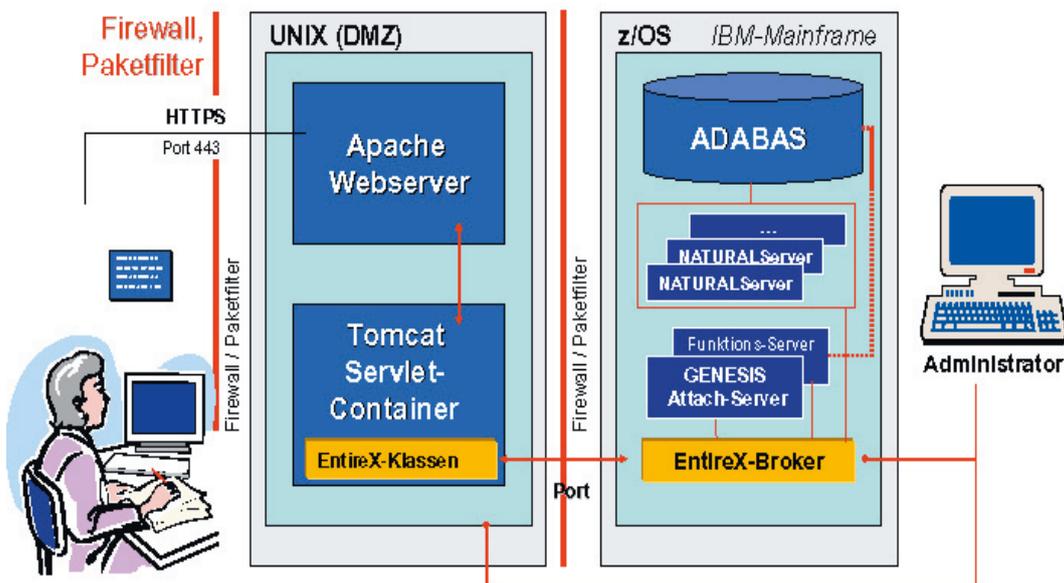
Die Tabelle kann in den Formaten MS-Excel®, HTML und CSV auf dem eigenen Rechner (PC) gespeichert werden.

Technik – Systemarchitektur der Datenbank

GENESIS-Online besteht aus

- dem GENESIS-Datenbankserver auf dem IBM-Großrechner des Landesamts, in dem die Daten gespeichert sind und die GENESIS-Anwendungslogik abgearbeitet wird und
- der Oberflächenkomponente als Schnittstelle zwischen dem Nutzer und dem Datenbankserver, bestehend aus einem Apache-Webserver und dem Tomcat Servlet-Container, der die vom Datenbankserver gelieferten Dialog- und Wertedaten mittels Java-Code und xml/xsl-Techniken aufbereitet. Diese Komponente läuft auf einem Unix-Rechner.

Datenbankkomponenten



Datenabruf –
technischer
Ablauf

Eine Nutzeranfrage, die über den Browser abgeschickt wird, wird an der Firewall geprüft und an den Webserver übergeben. Dieser startet ein Java-Servlet, welches die Datenaufbereitung (Formatkonvertierung) übernimmt. Mittels der Transportsoftware EntireX wird die Anfrage vom Server über eine zweite Firewall an den EntireX Broker des Großrechners übergeben. Dort erfolgt die Übergabe an den GENESIS - Datenserver. Die aus der Datenbank abgerufenen Informationen werden auf dem gleichen Weg an den Nutzer zurück gesandt.

GENESIS-Online (Bayern) – eine erste Nutzungsübersicht

GENESIS-
Online wird
gut ange-
nommen

Das Angebot von GENESIS-Online (Bayern) wird gut angenommen. Bislang wurden 2005 in rund 26 500 GENESIS-Sitzungen etwa 32 000 Werteabrufe getätigt, was einer Zahl von 2 650 Sitzun-

gen und 3 200 Werteabrufen je Monat entspricht. Vergleicht man die Abrufhäufigkeit monatsweise mit der des Vorjahres, so ergibt sich eine Steigerung um durchschnittlich 60 %.

Dabei werden zwei Drittel der Sitzungen über den Gastzugang aktiviert. Die durchschnittliche Verweildauer aller Nutzer im System beträgt rund 15 Minuten.

Bis Ende Oktober 2005 waren rund 4 000 Nutzer in GENESIS registriert, d.h. durchschnittlich werden ca. 180 Registrierungen im Monat vorgenommen. 96 % der registrierten Nutzer stammen aus Deutschland und verteilen sich hauptsächlich auf die Kategorien „Unternehmen“, „Politik und Verwaltung“, „Bildung/Wissenschaft/Kultur“ sowie „Privatpersonen“.