

Hans Dicken

Data Warehouses für die Forschungsstatistik

HIS: Forum Hochschule
2 | 2006

1. Einführung

In der öffentlichen Diskussion nehmen die Bereiche Forschung, Entwicklung und Innovation seit geraumer Zeit einen prominenten Platz ein. In der Tat verfolgt auch die neue Bundesregierung die im März 2000 vom EU-Rat beschlossene Lissabon-Strategie [1], wonach die Ausgaben der öffentlichen Gebietskörperschaften und der Wirtschaft einen Anteil von 3 Prozent des Bruttoinlandsprodukts erreichen sollen, was allein in Deutschland eine Steigerung der betreffenden Investitionen von weit mehr als 10 Milliarden Euro voraussetzt (im Jahre 2004 betrug die FuE-Ausgaben in Deutschland 55,6 Milliarden [2]).

Wie aber gelangt man eigentlich an diese Daten in der benötigten Form? Wie werden sie erhoben, Systematiken zugeordnet, aktualisiert und gepflegt? Tatsächlich basiert die Forschungsberichterstattung des Bundes auf einem relativ komplexen Informationsfluss von Haushalts-, Vorhabens-, Projekt- und anderen Daten, die aus sehr heterogenen Quellen zusammengeführt und verdichtet sowie um internationale Vergleichsdaten ergänzt werden. Gespeichert sind diese in einem von der HIS GmbH entwickelten und betreuten Data Warehouse mit zahlreichen unterschiedlichen Anwendungskomponenten, wobei die Daten teilweise aus operativen Systemen des BMBF (insbesondere die Vorhabens-basierte Datenbankanwendung PROFIL), teilweise aus im System integrierte Online-Formulare und teilweise aus weiteren externen Quellen gespeist werden. Wird etwa an einer Hochschule ein vom BMBF zu finanzierendes Forschungsvorhaben vom Bund bewilligt, dann wird der/die Projektverantwortliche mit einem sogenannten easy-Formular [3] konfrontiert, in das die Eckdaten des Vorhabens einzutragen sind (u. a. Beschreibung des Vorhabens, Personalausstattung, Laufzeit, Finanzierungsplan). Dieses Formular wird in der Regel online ausgefüllt, wodurch die Einträge direkt in die operative Vorhabensdatenbank PROFIL gelangen, die im Rechenzentrum des Deutschen Zentrums für Luft- und Raumfahrt (DLR) in Köln lokalisiert ist. Selbstverständlich sind nach Einzelvorhaben gegliederte Ausgaben für Veröffentlichungen nicht geeignet, sie werden daher mittels stichtagsbezogener Snapshots in eine gesonderte Datenbank namens DASTAT auf eine Haushaltsebene verdichtet (zumeist auf die Ebene von Haushaltstiteln) und sind dort auswertbar. DASTAT beinhaltet ebenfalls die Soll- und Plandaten des Bundeshaushalts und unterschiedliche Untergliederungen (regionale, inhaltliche und weitere), deren Zuordnungen entweder durch automatische Mechanismen aus PROFIL oder direkt durch die Bearbeitung mittels spezieller DASTAT-Anwendungen im BMBF geschehen. Die Struktur des Bundeshaushalts wird einer weiteren Datenbank namens IT-Haushalt entnommen und mit den Daten aus PROFIL und den anderen Quellen gefüllt, seit kurzem ist IT-Haushalt (in technischer Hinsicht) in PROFIL aufgegangen.

DASTAT-Auswertungen bilden einen Teilbereich der für die periodische FuE-Berichterstattung des Bundes benötigten Daten ab (nämlich die Ausgaben des Bundes und der Länder), um aber beispielsweise den Bundesbericht Forschung oder Faktenbericht Forschung, bzw. die entsprechenden Internet-Veröffentlichungen des BMBF (z.B. [2]) zu erzeugen, werden zusätzliche Daten benötigt, darunter die Ausgaben der Wirtschaft, internationale Vergleichsdaten und Indikatoren. Diese Daten sind in einer weiteren Datenbank namens FOSTAT organisiert, aus der direkt Tabellen und Abbildungen in einem vollkommen fertigen Layout mit aktuellen Daten abgerufen werden können, die dynamisch auf Abruf erzeugt werden.

DASTAT (in der aktuellen Form) und FOSTAT sind Entwicklungen der HIS GmbH. In den folgenden Abschnitten werden beide Anwendungen in inhaltlicher und technischer Hinsicht vorgestellt. Daneben wird die Organisation der Abläufe bei Auswertungen, Datenaktualisierungen,

inhaltlichen Pflegearbeiten, usw. beleuchtet. Insbesondere FOSTAT ist nicht auf die gegebenen Inhalte beschränkt oder speziell zugeschnitten, man könnte dieses Data Warehouse ebenfalls beispielsweise für vergleichbare Aufgaben auf Länder- oder Hochschulebene einsetzen.

2. DASTAT: Abbild des Bundeshaushalts im Bereich FuE

DASTAT ist originär eine komplexe Datenbankanwendung zur Verwaltung und Auswertung der Ausgaben des Bundes für Wissenschaft, Forschung und Entwicklung mit dem BMBF als Hauptanwender und Auftraggeber. Auf Anforderung des Ministeriums wurde die Datenbasis vor einigen Jahren um die Forschungsausgaben der Länder und der Bildungsausgaben des BMBF erweitert. Letzteres führt dazu, dass mit Hilfe von DASTAT nahezu der gesamte Haushalt des BMBF abgebildet werden kann (eine Ausnahme bilden lediglich die Verwaltungsausgaben des Ministeriums). Damit hat man gleichzeitig die Möglichkeit, diese Daten in dem Sinne zu verknüpfen, dass man beispielsweise nach bildungsrelevanten FuE-Ausgaben fragt oder nach Wissenschaftsausgaben, die nicht FuE sind.

Im folgenden Abschnitt wird zunächst eine grobe terminologische und inhaltliche Abgrenzung versucht, anschließend werden die Datenflüsse nach DASTAT und von DASTAT nebst inhaltlicher Untergliederungen dargelegt, bevor kurz auf die Herkunft und Verwaltung der Länderausgaben eingegangen wird. Am Ende des Kapitels wird die Anwendung in organisatorischer und technischer Hinsicht (mit einer ausführlicheren Darstellung der Benutzersicht) beschrieben.

2.1 Begriffliche und inhaltliche Abgrenzungen

Beschäftigt man sich mit Ausgaben für Wissenschaft, Forschung und Entwicklung, dann wird man zunächst nach den Terminologien selber fragen. Das Frascati Manual der OECD [4] beantwortet Fragen nach FuE auf der begrifflichen Ebene recht erschöpfend: Forschung und experimentelle Entwicklung (FuE) ist demnach die systematische, schöpferische Arbeit zur Erweiterung des vorhandenen Wissens einschließlich des Wissens über den Menschen, die Kultur und die Gesellschaft sowie die Verwendung dieses Wissens mit dem Ziel, neue Anwendungsmöglichkeiten zu finden (Frascati Manual 1993, § 57).

Auf der Ausgabenseite sind damit allerdings keineswegs alle Fragen beantwortet. Sind z.B. die Aufwendungen für Machbarkeitsstudien von FuE-Projekten Forschungsausgaben? In der Tat begegnet man derartigen Fragen auf der Anwenderseite außerhalb des BMBF relativ häufig. Ein Auszug aus einem Hilfetext von DASTAT (inhaltlich formuliert vom BMBF) illustriert, wie detailliert die Vorgaben bei der Zuordnung sind (bzw. sein müssen):

„Insbesondere im Rahmen der institutionellen Förderung sowie in Bundesanstalten mit Forschungsaufgaben sind FuE und forschungsverwandte Tätigkeiten oft so eng miteinander verbunden, dass eine exakte getrennte Nachweisung der FuE-Ausgaben schwierig ist. In diesen Fällen ist der FuE-Anteil eines Titels bzw. Kapitels sorgfältig zu schätzen, wobei folgende Tätigkeiten der FuE wertmäßig zuzurechnen sind:

- Personalausgaben, sächliche Verwaltungsausgaben sowie Bau- und sonstige Investitionsausgaben, die unmittelbar mit intern durchgeführter Forschung und Entwicklung in Verbindung stehen sowie anteilige Ausgaben für die Infrastruktur, z.B. für Verwaltungspersonal.
- kalkulatorische Ausgaben, die der Nutzung eigener Einrichtungen durch Dritte für FuE-Zwecke entsprechen.
- Ausgaben für Forschungsaufträge an Dritte sowie die kalkulatorischen Ausgaben für das Personal, das mit der Vergabe von FuE-Aufträgen an Dritte beschäftigt ist.
- Zuschüsse für Forschungszwecke an Dritte.“

Wissenschaftsausgaben umfassen demgegenüber Ausgaben für Forschung und Entwicklung (FuE) sowie Ausgaben für wissenschaftliche Lehre und Ausbildung und sonstige verwandte wissenschaftliche und technologische Tätigkeiten. Zu letzteren gehören z. B. wissenschaftliche und technische Informationsdienste, Datensammlung für allgemeine Zwecke, Untersuchungen über die Durchführbarkeit technischer Projekte (Durchführbarkeitsstudien von Forschungsvorhaben sind jedoch Teil von FuE), Erarbeiten von Grundlagen für Entscheidungshilfen für Politik und Wirtschaft.

DASTAT unterscheidet zudem bildungsrelevante FuE-Ausgaben (z.B. für die Hochschulforschung) und bildungsrelevante Wissenschaftsausgaben, die nicht gleichzeitig FuE-Ausgaben sind (z.B. Hochschulbau). Eine weitere wichtige Definition betrifft die Grundlagenforschung (GLF): Diese ist definiert als experimentelle oder theoretische Arbeit, die in erster Linie auf die Gewinnung neuer Erkenntnisse über den zugrundeliegenden Ursprung von Phänomenen und beobachtbaren Tatsachen gerichtet ist, ohne auf eine besondere Anwendung oder Verwendung abzuführen (vgl. Frascati Manual 1993, § 224). Neben der „reinen Grundlagenforschung“, die zur Mehrung des Wissens ausgeführt wird, ohne dass sich langfristig wirtschaftlicher oder sozialer Nutzen erkennen lässt, ist auch „anwendungsorientierte Grundlagenforschung“ zu berücksichtigen, die ausgeführt wird mit der Erwartung, dass sie langfristig eine breite Wissensgrundlage schafft, die als Ausgangsbasis für die Lösung erkannter oder erwarteter, bestehender oder künftiger Probleme oder Entwicklungen dient, z.B. im Rahmen der Materialforschung, Informationstechnik oder Biotechnologie.

Für weitergehend Interessierte findet sich eine Auflistung der „Frascati Family“ of Guidelines for the Measurement of Scientific and Technological Activities im Oslo Manual [5].

2.2 Von den operativen Systemen des BMBF zu DASTAT

Wie oben angedeutet, bezieht DASTAT seine Daten im wesentlichen aus drei Quellen:

- Den operativen Systemen des BMBF
- Direkten Online-Eingaben in den Bundesressorts und nachgeordneten Behörden
- Quasi-direkten Dateneingaben des Statistischen Bundesamts (Länderausgaben)

Die Behandlung der FuE-Ausgaben der Länder über das Statische Bundesamt und die direkten Online-Eingaben der Ressorts werden in späteren Abschnitten etwas ausführlicher dargestellt, der Großteil der FuE-Ausgaben werden aus den operativen Systemen des BMBF nach DASTAT übernommen. Insbesondere die Ausgaben des BMBF (knapp 70% der FuE-Ausgaben des Bundes bis vor der Ressortumstrukturierung infolge der Regierungsneubildung 2005) verwaltet alle seine Projekte über die operative Vorhabensdatenbank PROFi (die vom BMBF in Zusammenarbeit mit der

DLR betrieben wird). Auch einige andere Ressorts benutzen teilweise bereits PROFi, mittelfristig wird daran gedacht, sämtliche Bundesministerien an dieses Online-System anzuschließen.

PROFI ist ein operatives System in dem Sinne, dass dessen Inhalte im Gegensatz zu denen in DASTAT vergleichsweise erheblich dynamischer und flüchtiger sind. Es können täglich einzelne Vorhaben hinzukommen, sich ändern oder verschwinden (etwa weil sie abgeschlossen sind). Für Zwecke des Berichtswesens, von Anfragen aus dem Parlament, bzw. der Öffentlichkeit oder zur Lieferung von nationalen Vergleichsdaten an internationale Institutionen werden demgegenüber Daten benötigt, die eine gewisse zeitliche Abgeschlossenheit und Konsistenz aufweisen. Aus diesem Grund werden PROFi-Daten nur zu festgelegten „geeigneten“ Zeitpunkten als zeitliche Snapshots in verdichteter Form nach DASTAT übernommen. Ein „geeigneter Zeitpunkt“ ist zumeist dann gegeben, wenn die IST-Ausgaben des Bundes für Wissenschaft und FuE bezogen auf ein Haushaltsjahr vollständig vorliegen, dies ist gewöhnlich im Frühjahr des folgenden Jahres gegeben.

Vor den eigentlichen Datenübernahmen (IST-Zahlen des Vorjahres) erfolgt eine Aktualisierung der Haushaltsstruktur des Bundes (und die Übernahme von SOLL- und PLAN-Daten aus in die Zukunft gerichteten Haushaltsplänen und aus der mittelfristigen Finanzplanung). Bislang wurden die dafür benötigten Daten in einer eigenen Datenbank verwaltet, mittlerweile ist auch diese in PROFi aufgegangen.

Die einzelnen Projekte aus PROFi werden im Laufe der Datenübernahmen DASTAT-Haushaltstiteln zugeordnet, ein solcher Titel ist z.B. „Meeres- und Polarforschung, Geowissenschaften“. Anschließend werden diese anteilig auf unterschiedliche Systematiken verteilt, mit geographischen Merkmalen versehen, etc. um Auswertungen nach unterschiedlichen Merkmalen vornehmen zu können.

2.3 Stammdaten, Ausgaben und inhaltliche Zuordnungen von FuE-Titeln

Ein DASTAT-Titel besteht grob aus drei Bestandteilen:

- Titelstammdaten
- Zuordnungen/Unterteilungen zu Systematiken und anderen Merkmalen
- Ausgaben (Beträge)

Die Titelstammdaten beinhalten Metainformationen zum jeweiligen Titel insgesamt, darunter numerische Schlüssel zur Identifizierung, eine textuelle Beschreibung, die Laufzeit, eine Förderungsart (z.B. „Direkte Projektförderung“, „Hochschul- und bildungsbezogene Förderung“). Abbildung 1 zeigt eine DASTAT-Pflegemaske mit geöffneten Stammdaten zu einem Titel, teilweise sind diese tatsächlich nur für Fachanwender verständlich.

Die Ausgaben sind für jeden Titel in die folgenden Kategorien unterteilt:

- Gesamtausgaben
- Länderaufteilung (bezogen auf die ausführende Stelle)
- Aufteilung nach Empfängergruppen (z.B. Großforschungseinrichtungen)
- Aufteilung nach Einrichtungen (oftmals einzelne Institute)
- Aufteilung nach Wirtschaftszweigen (z.B. Maschinenbau)

The screenshot shows the DASTAT software interface for entering title data. At the top, there is a menu bar with options like 'Titel kopieren', 'Hilfe', and 'Excel'. Below the menu is a header section with fields for 'Stand T_Nr', 'Länderkennung', 'Finanz.art', 'Einzelplan', 'Kapitel', 'Titel', 'Teilwert', 'Zuordnung', 'Fkt', and 'Laufzeit'. The main area is divided into several tabs: 'Titelstammdaten', 'Systematiken', 'Ausgaben', 'Titelstruktur', 'Einrichtungen', 'Abkürzungstabellen', 'Freigabe', and 'Protokollierung'. The 'Titelstammdaten' tab is active, showing fields for 'Zweckbestimmung-Langfassung', 'Zweckbestimmung-Kurzfassung', 'Titelreferat', 'GFE', 'Förderart', and 'Notizen zum Titel'. The 'Einrichtungen' tab is also visible, showing a table for 'Ausgaben zu Einrichtungen für das Jahr' with columns for 'DS_E...', 'Jahr', 'Typ', 'Einrichtung', and 'T€'. Below this, there is a section for 'Automatische Verknüpfungen von Titeln mit identischer LP- und NABS-Zuordnung' and a table for 'Einrichtungsnr.' with columns for 'Bundesl...', 'T_NR', 'Kapi...', 'Titel', 'Teilwert', and 'verknü.'.

Abb. 1: DASTAT-Titelstammdaten

Daneben werden zahlreiche weitere Zuordnungen und Systematiken in der Datenbank verwaltet, z. B. der Anteil an Grundlagenforschung, Zuordnungen zur Leistungsplansystematik des Bundes und der NABS-Systematik der EU (Nomenclature pour l'analyse et la comparaison des budgets et programmes scientifiques), einem Ausgabentyp (IST, SOLL oder PLAN), geographische Zuordnungen und Informationen darüber, ob es sich um Ausgaben an Entwicklungsländer oder für internationale Zusammenarbeit handelt.

All die genannten Informationen können bei DASTAT-Auswertungen abgerufen werden.

2.4 FuE-Ausgaben der Länder

Die FuE-Ausgaben der Länder werden von den statistischen Landesämtern erhoben und anschließend an das Statistische Bundesamt übermittelt. Das StBA ist dann in der Lage, diese mittels eines speziell angepaßten DASTAT-Pflegemoduls in die Datenbank zu übertragen. Die Ausgangsdaten haben dabei teilweise andere Verschlüsselungssystematiken als die Haushaltsdaten des Bundes, was zu einem relativ hohen Umsetzungsaufwand führt (so werden Länderausgaben in der Regel sogenannten „Sektoren“ zugeordnet, was auf der Bundesebene „Empfängergruppen“ entspricht).

Länderausgaben in DASTAT werden im wesentlichen untergliedert nach der Länderzuordnung (finanzierendes Land), einer NABS-Gliederung, einer Empfängergruppenzuordnung (z.B. Forschungseinrichtungen der Länder) und einer Funktion des Forschungstitels (z.B. Denkmalpflege). Abbildung 2 zeigt das DASTAT-Pflegemodul für die FuE-Ausgaben der Länder. Da alljährlich weit mehr als 1000 Ländertitel einzeln zu aktualisieren und zu pflegen sind, werden derzeit alternative Möglichkeiten gesucht, die eine weniger zeitaufwändige Verarbeitung ermöglichen.

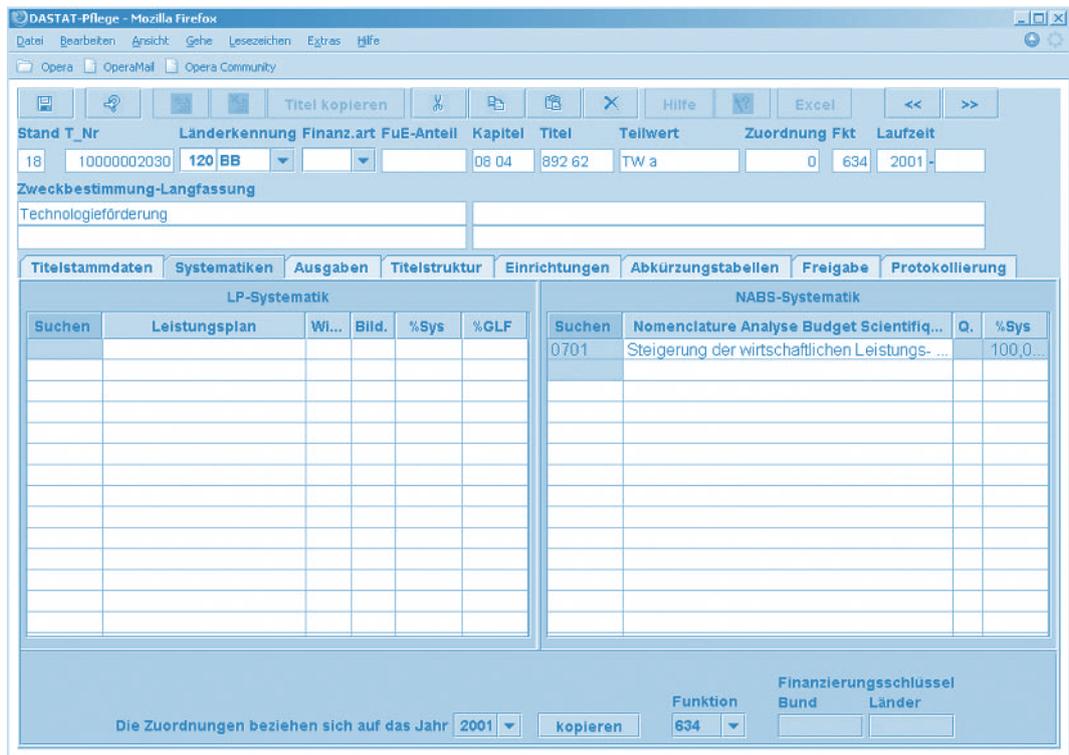


Abb2. Pflegemodul für FuE-Ausgaben der Länder

2.5 Datenkonsistenz trotz zeitlicher Dynamik

DASTAT beinhaltet Zeitreihen über Ausgaben für Wissenschaft, Forschung und Entwicklung, die bis in das Jahr 1974 zurückreichen. Innerhalb der Zeitspanne von diesem Jahr bis in die Gegenwart waren und sind die Systematiken für die Zuordnung der Daten aber sehr im Fluss. Man mag sich daher die Frage stellen, welchen Sinn Datenauswertungen haben, welche z. B. soweit zurückreichen, dass sie Ausgaben des längst aufgelösten Bundesministeriums für Post und Telekommunikation beinhalten. Die DASTAT-Datenbasis ist robust bezogen auf derartige strukturelle Modifikationen: Ändern sich z.B. die Zuschnitte einzelner Ressorts oder die Aufteilungen von Haushaltsstrukturen bezogen auf andere Untergliederungen, dann werden die geänderten Zuordnungen in DASTAT nachvollzogen, so dass jederzeit eine konsistente Gesamtsicht auf die Daten vorliegt.

Dennoch bleiben auch vergangene Zahlen und Zuordnungen vollständig erhalten. Dies wird durch zwei Mechanismen erreicht: Erstens sind die Daten in sogenannten „Ständen“ organisiert, wobei ein Stand immer einem Stichtag zugeordnet ist (der älteste Stand 1 trägt z.B. das Datum 30.03.1994), wobei ein neuer Stand initial immer als exakte Kopie des jeweils vorherigen angelegt wird, anschließend wird nur noch der neue Stand modifiziert, der alte bleibt als eine Art Archiv erhalten. Bei Auswertungen besteht der erste Schritt dann immer darin, anzugeben, in welchem Stand man auswerten will. Auch wenn dies in der Regel der aktuellste ist, kann man dennoch auch historische Daten abrufen.

Zweitens sind alle Titel mit einer Laufzeit versehen. Auslaufende Haushaltsstellen können damit in zukünftige Stände mitgeführt werden, ohne dass sie bei Aggregationen oder Summierungen stören. Selbstverständlich führt dies dazu, dass die Quantität der DASTAT-Daten perma-

ment zunimmt, dies findet aber auf einem relativ niedrigen Niveau statt, so dass damit einhergehende technische oder organisatorische Probleme in absehbarer Zeit nicht auftreten werden.

Diese Form der Datenorganisation geht mit einem relativ hohen Pflegeaufwand einher. In der Tat werden die Daten mit Hilfe von speziellen DASTAT-eigenen Softwaremodulen zumeist direkt durch das Bundesministerium für Bildung und Forschung kontrolliert und gepflegt.

2.6 Anwendungskomponenten

DASTAT-Anwendungswerkzeuge existieren insbesondere für Auswertungen in mehreren Varianten, für die Datenpflege und -aktualisierung, für Datenerhebungen über das Internet, bzw. über das Intranet des Bundes IVBB sowie für mehrere sonstige administrative Zwecke. In diesem Abschnitt sollen die Auswertungskomponente und die (organisatorisch und somit vom technischen Aufwand her) komplexe Anwendung zur Datenerhebung in den Ressorts und deren nachgeordneten Behörden vorgestellt werden, aus Platzgründen in sehr gestraffter Form.

2.6.1 Auswertungen

DASTAT-Auswertungskomponenten existieren in mehreren Varianten, wobei sich die Funktionalität im Kern immer ähnelt; die Unterscheidungen resultieren vielmehr aus der unterschiedlichen Sichtbarkeit von Daten für unterschiedliche Anwenderkreise: Jeder Anwender sieht nur die für ihn relevanten Daten, einzig das BMBF als verantwortliches Ressort hat Zugang zum gesamten Datenspektrum.

Nach der Anmeldung und der Auswahl des Standes, in dem ausgewählt wird, sieht der Benutzer die in Abbildung 3 dargestellte Maske.

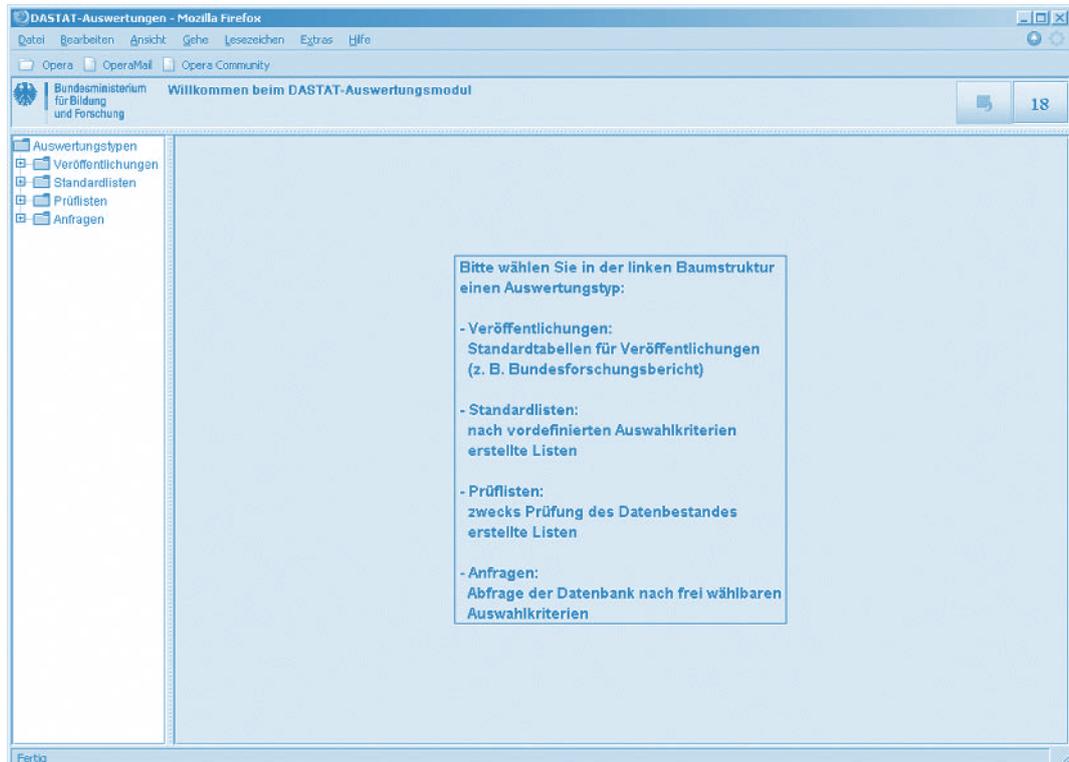


Abb. 3 Ausgangsdialog bei Anwendungen

Der Benutzer/die Benutzerin kann an dieser Stelle zwischen vier verschiedenen Typen von Auswertungen auswählen: Den strukturell vordefinierten Standardveröffentlichungstabellen (dazu zählen unter anderem Tabellen über die Ausgaben des Bundes nach Empfängergruppen, nach finanzierenden Ressorts oder nach Förderarten), den vornehmlich für BMBF-interne Zwecke relevanten Standardlisten und Prüflisten sowie den flexiblen Abfragen nach frei wählbaren Auswahlkriterien. Wird nun die hinter dem Wörtchen „Anfragen“ steckende Baumstruktur aufgeklappt, dann öffnen sich mehrere sogenannte Anfragekategorien (z.B. Gesamtvolumen von Titeln, Untergliederungen nach Bundesländern oder Empfängergruppen oder auch spezielle Ausgabenarten). Die meisten dieser Kategorien kann man wiederum aufklappen, um zu immer feiner gegliederten Daten zu gelangen. Abbildung 4 zeigt den Dialog, nachdem die Kategorie „Gesamtvolumen“ ausgewählt wurde.

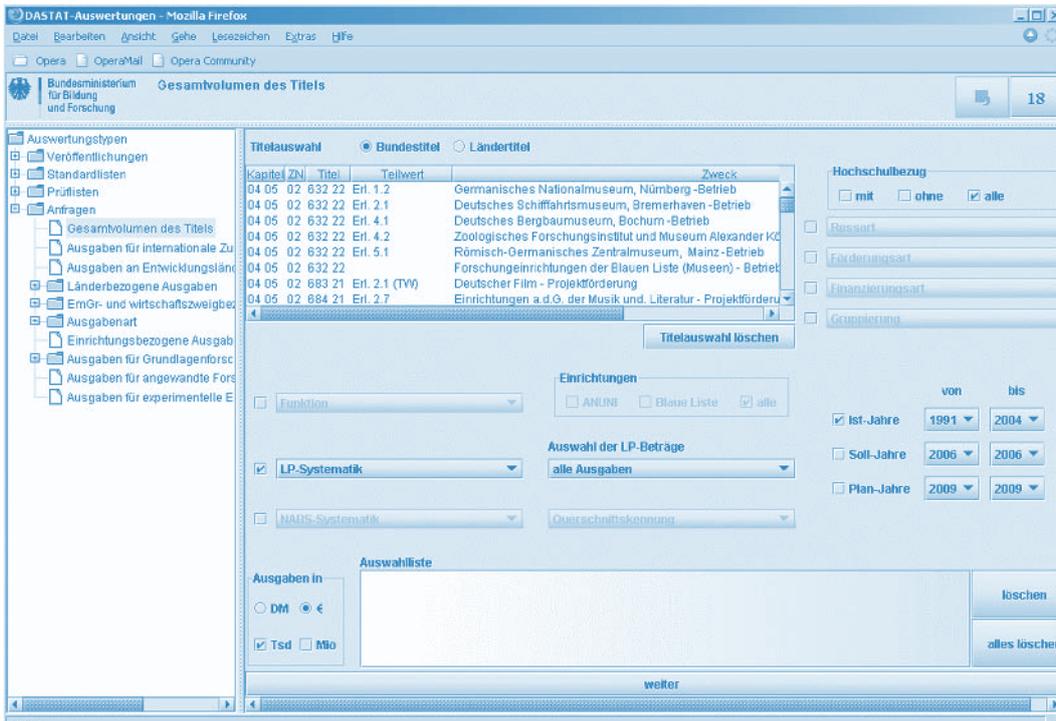


Abb. 4 Auswahldialog für Titel-Gesamtvolumen

Innerhalb dieser Maske nun besteht die Möglichkeit, die Auswahl der auszuwertenden Ausgaben nach verschiedenen Kriterien einzuschränken, wobei die Möglichkeiten von der Auswahl eines Einzeltitels über diverse weniger restriktive Einschränkungen bis hin zum Einbeziehen aller Haushaltstitel reichen. Durch entsprechende Einschränkungen des Ressorts und der Förderungsart könnte man sich somit etwa die Ausgaben des Bundesumweltministeriums für direkte Projektförderung ausgeben lassen, wobei das System zunächst alle in Frage kommenden Titel weiter verarbeitet und in späteren Schritten abhängig vom Grad der Einschränkungen automatisch Summen bildet.

Nachdem diese grundlegende Definitionen vorgenommen sind, kann im nächsten Schritt das Aussehen der Ergebnistabelle festgelegt werden (Abbildung 5), bevor das gewünschte Ergebnis angezeigt wird (Abbildung 6).

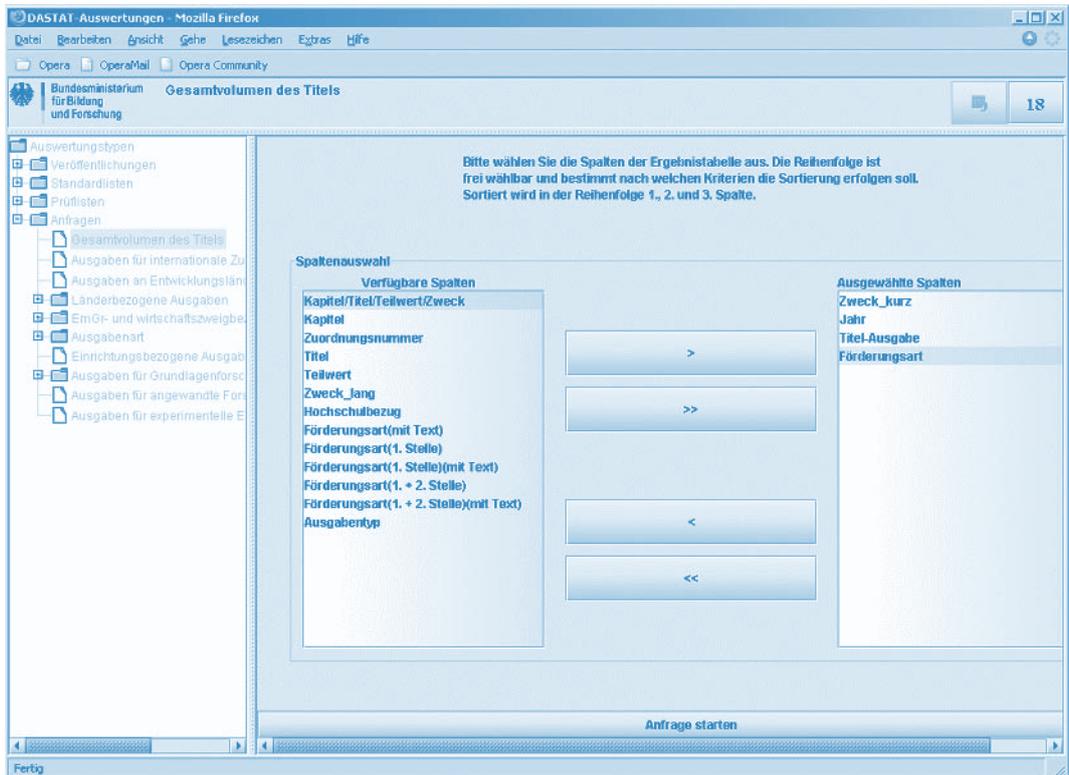


Abb. 5: Festlegen des Aussehens der Ergebnistabelle

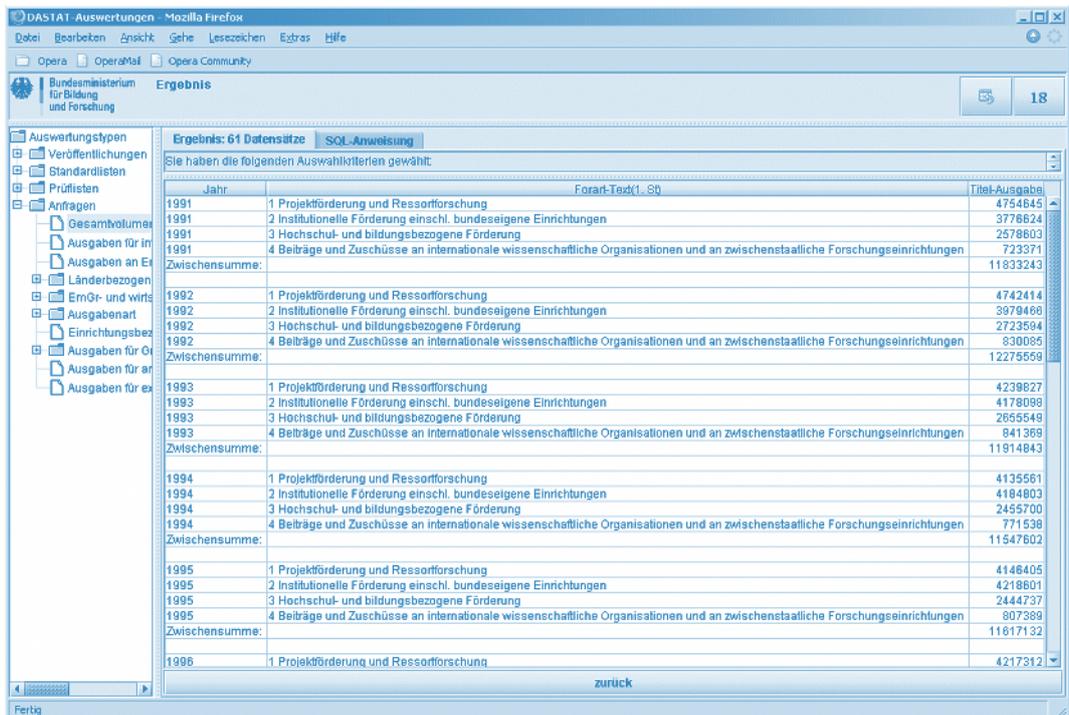


Abb. 6: DASTAT-Anfrageergebnis

Das Ergebnis der Anfrage kann nun mittels unterschiedlicher Operationen sortiert oder summiert werden (wobei auch Zwischensummen gebildet werden können), zur weiteren externen Verarbeitung ist zudem auch die Konvertierung und Abspeicherung im Microsoft Excel-Format möglich.

Nach diesem Muster sind in DASTAT sehr vielfältige ad hoc-Anfragen möglich, es konnte hier lediglich ein kleiner Ausschnitt der Möglichkeiten gezeigt werden.

2.6.2 *Online-Erhebungen in den Ressorts und nachgeordneter Bundesbehörden*

Während die Ausgaben des BMBF für Wissenschaft, Forschung und Entwicklung bei anstehenden Aktualisierungen direkt aus der operativen PROFI-Datenbank übernommen und anschließend an DASTAT-Bedürfnisse angepaßt werden, verwalten die meisten anderen Bundesressorts ihre Daten in unterschiedlicher Form selber. Eine Vielzahl von Titeln werden zudem von nachgeordneten Bundesbehörden verwaltet, die einzelnen Ressorts zugeordnet sind. Allen Ressorts gemeinsam ist, dass pro Ministerium ein Forschungsbeauftragter existiert, der die Dateneingaben intern initiiert und kontrolliert. Das BMBF wiederum meldet den Forschungsbeauftragten, dass Titel zur Erhebung anstehen und nimmt die erhobenen Daten in Produktionsbetrieb (d.h. erst nach der Kontrolle der Daten durch den jeweiligen Forschungsbeauftragten und durch das BMBF sind die betreffenden Daten auswertbar, vorher befinden sie sich in einem speziell für die Erhebungen eingerichteten temporären Datenbanksegment).

All diese hierarchischen Vorgänge finden online mit Hilfe von DASTAT-Modulen statt, daneben auch zusätzliche administrative Arbeiten der Forschungsbeauftragten (Verwaltung der Mitarbeiter, Zuordnung von Titeln zu Mitarbeitern). Abbildung 7 zeigt den strukturellen Aufbau.

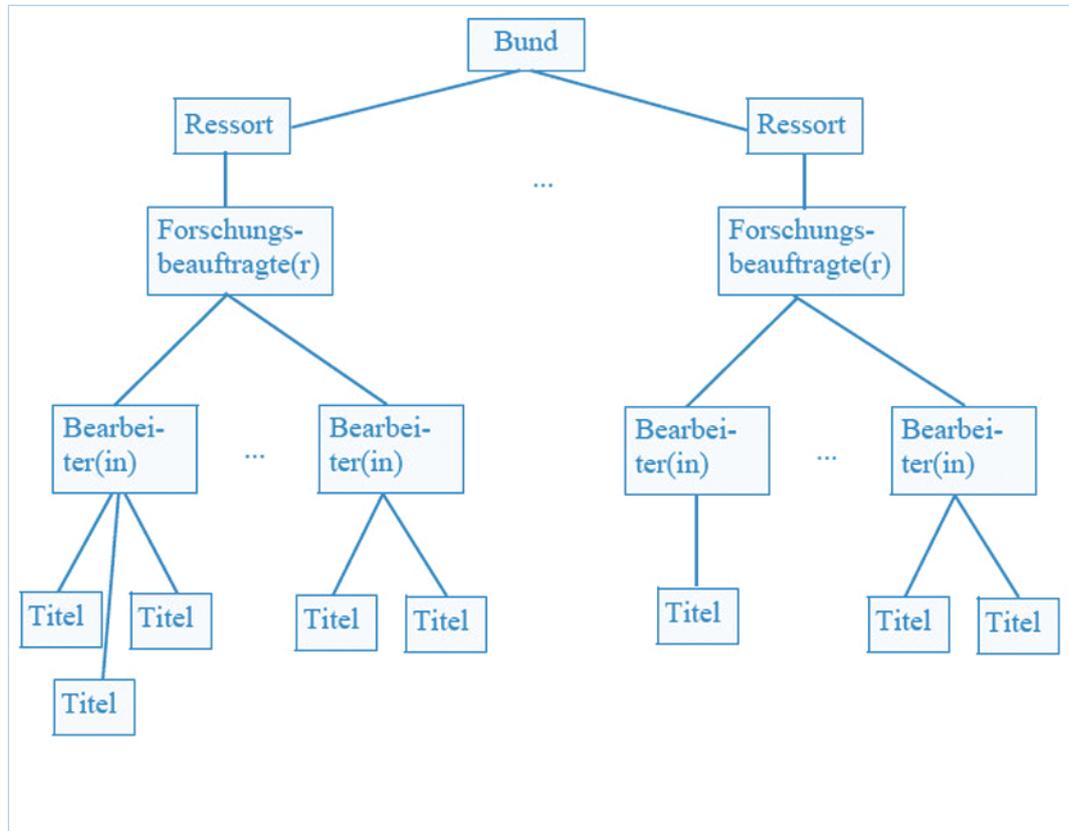


Abb. 7: Organisation und Zuordnung von Titeln zu Personen

Abbildung 8 zeigt den Ablauf der jährlichen Datenerhebungen in (etwas vereinfachter) schematischer Form: Sobald die Eingaben in den Ressorts anstehen, wird im BMBF die aktuelle Haushaltsstruktur in einen temporären Bereich der Datenbank kopiert, womit der jeweils relevante Ausschnitt für die Forschungsbeauftragten sichtbar werden. Diese ordnen den einzelnen Haushaltsstellen zuständige Bearbeiter ihres Ressorts zu (falls diese im Vergleich zum Vorjahr nicht unverändert sind). Damit sind die betreffenden Titel für die Bearbeiter sichtbar und können mit aktuellen Daten und Zuordnungen versehen werden. Nach Prüfung der Eingaben auf Plausibilität und Korrektheit übermittelt der Forschungsbeauftragte diese wieder an das BMBF, welche sie vom temporären Segment in die für Auswertungen zugängliche Datenbank transferiert.

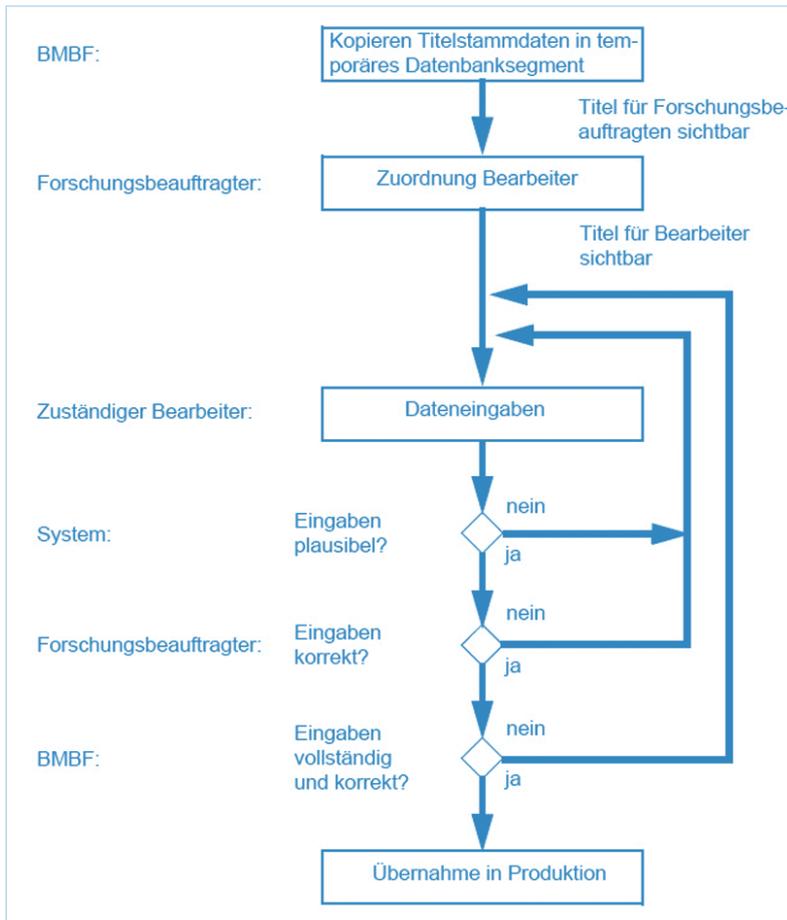


Abb. 8: Ablauf der DASTAT-Ressorthebungen

2.7 Technische Architektur im Überblick

Gute Internetanwendungen zeichnen sich unter anderem dadurch aus, dass möglichst geringe Anforderungen an die Rechner des Benutzers zu stellen sind (z.B. [6]). Während man bis vor einigen Jahren für Datenbankanwendungen die jeweilige Bearbeitungssoftware auf dem Rechner eines jeden Benutzers zu installieren hatte, welche dann lediglich noch die zu verarbeitenden Daten von einem Server lud (sogenanntes Client/Server-Prinzip), sind moderne Internet/Intranet-Anwendungen so ausgelegt, dass die Verarbeitungssoftware weitgehend auf einem weiteren Server lokalisiert ist und dass der Anwender lediglich noch Anfragen an den Applikationsserver transferiert, welcher seinerseits dann mit dem Datenbankserver kommuniziert, um die Anfrageergebnisse zu ermitteln und an den Benutzer zurückzuschicken (Multi-Tier-Prinzip).

Eine geeignete Arbeitsumgebung für solche Architekturen liefert Java: Man kann (anders als bei einfachen Web-Anwendungen) funktional bequem zu handhabende und ergonomische Oberflächen in Form von Applets erzeugen und auf der Serverseite bietet jede neuere Entwicklung eine Java-Schnittstelle für die Kommunikation mit dem Benutzer auf der einen Seite und mit der Datenbank auf der anderen Seite. Die DASTAT-Anwendung arbeitet nach diesem Prinzip, als Datenbanksystem ist Oracle in Benutzung. Eine nähere technische Beschreibung findet sich in [7].

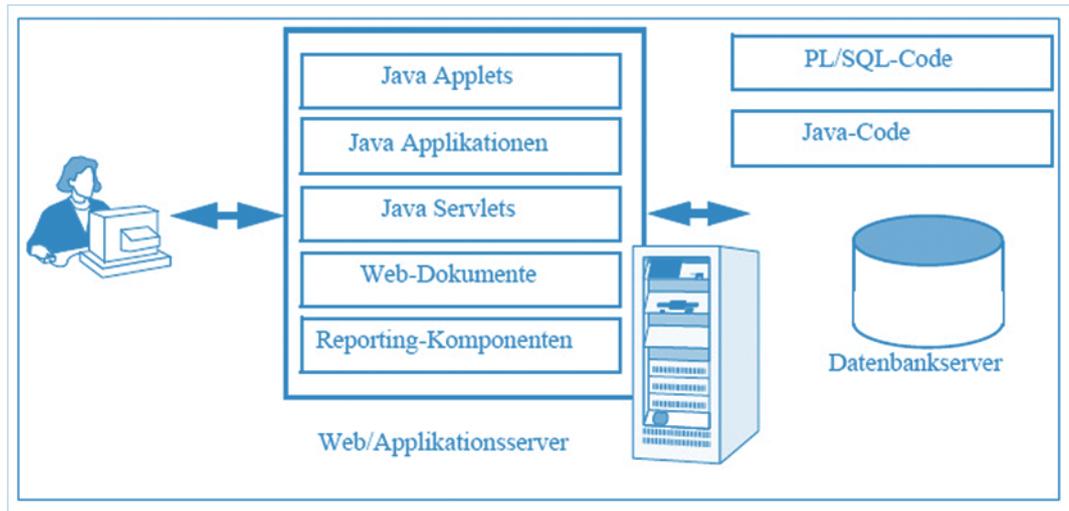


Abb. 9: Mehrschichtarchitektur in DASTAT

Für interne IT-Anwendungen der Bundesministerien wurde der Informationsverbund Berlin-Bonn (IVBB) eingerichtet, das ist ein Intranet des Bundes, innerhalb dessen wiederum einzelne Intranets für die Ressorts angelegt sind. Im IVBB finden äußerst rege interne (sichere) Informationsaustausche statt, das Hauptmerkmal des Bundesintranet besteht darin, dass Verbindungen von innen nach außen technisch zugelassen sind, von außen nach innen dagegen restriktiv nicht. Ein großer Teil der DASTAT-Anwender befindet sich innerhalb des IVBB, für diese wurde (optional) ein vergleichsweise sehr schneller Intranet-Zugang zu einem Server innerhalb des BMBF eingerichtet. Für Anwender außerhalb des IVBB besteht zudem die Möglichkeit, auf die Anwendungen über einen in der HIS GmbH lokalisierten Server zuzugreifen; beide Anwendungsserver greifen auf die selbe Datenbank zu, welche beim Deutschen Zentrum für Luft- und Raumfahrt (DLR) in Köln angesiedelt ist. Sämtliche Verbindungen sind verschlüsselt.

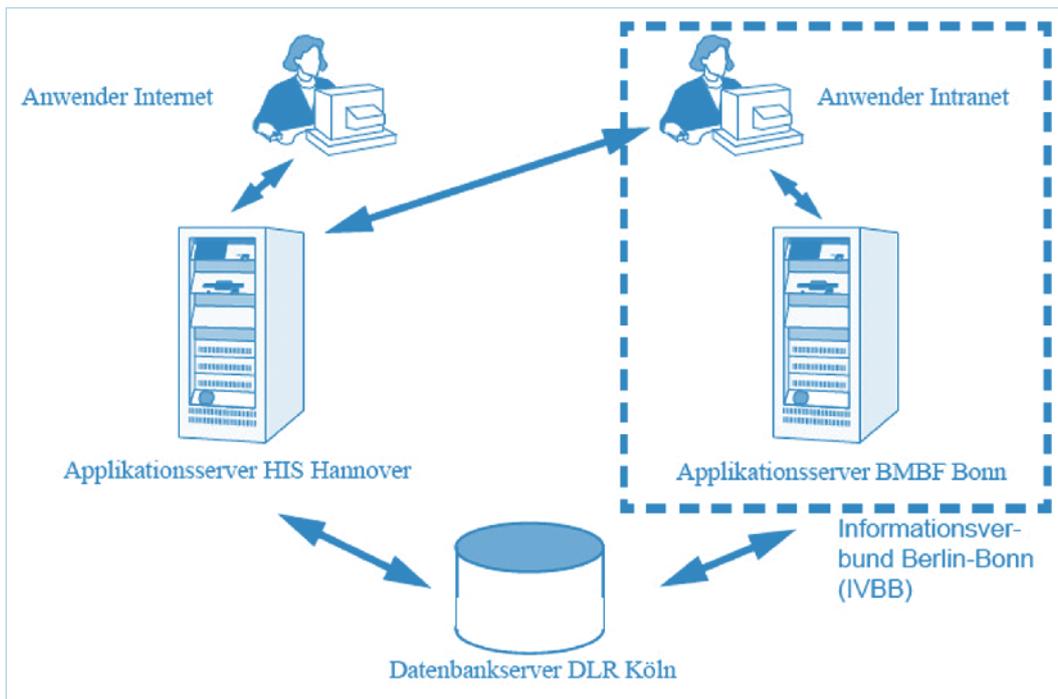


Abb. 10: DASTAT-Standorte

3. FOSTAT: Unterstützung der Forschungsberichterstattung

DASTAT beinhaltet im Kern Ausgaben des Bundes für Wissenschaft, Forschung und Entwicklung, daneben Informationen über FuE-Ausgaben der Länder und Bildungsausgaben des BMBF. Für die periodische Forschungsberichterstattung des Bundes aber wird auf der einen Seite derzeit nur ein kleiner Ausschnitt dieser Daten benötigt, andererseits beinhalten die Veröffentlichungen zahlreiche weitere Fakten, beispielsweise die Ausgaben der Wirtschaft, internationale Vergleichsdaten und nicht-monetäre Indikatoren.

Diese werden in einer weiteren Datenbankanwendung mit der Bezeichnung FOSTAT verwaltet. FOSTAT beinhaltet nicht nur die Daten selber, sondern zusätzlich dezidierte Layout-Informationen für Tabellen und Abbildungen, die auf Knopfdruck mit jeweils aktuellen Daten in einem publikationsfertigen Dateiformat abgerufen werden können.

3.1 Veröffentlichungen des Bundes im Bereich Forschung und Entwicklung

Zweijährlich alternierend erschienen bislang der Bundesbericht Forschung und der Faktenbericht Forschung, der letzte Bundesbericht im Jahre 2004 ([8]). In mehr als 60 Tabellen, zahlreichen Abbildungen und umfangreichen textuellen Erläuterungen werden in diesen Berichten aktuelle Fakten und Übersichten zur deutschen Forschungslandschaft zusammengefaßt und erklärt. Die Tabellen und Statistiken befassen sich aus einer etwas vergrößerten Sicht mit den folgenden Themenkomplexen:

- Ausgaben für Wissenschaft, Forschung und Entwicklung
- FuE-Personal
- Indikatoren zur technologischen Leistungsfähigkeit
- FuE-relevante Bildungsdaten
- Ausgewählte Kennzahlen zur wirtschaftlichen Entwicklung

Die gedruckten Veröffentlichungen wurden im Jahre 2005 um eine Internet-Veröffentlichung sämtlicher Tabellen und einiger ausgewählter Abbildungen ergänzt ([2]). Die Internet-Version bildet (bezogen auf die Vergangenheit) in vielen Fällen einen längeren Zeitraum ab, als die gedruckten Veröffentlichungen. In der Datenbank allerdings befindet sich noch ein weitaus größerer Bestand an Daten, dieser wird in Zukunft die Basis der Veröffentlichungen des BMBF im Bereich Wissenschaft, Forschung, Entwicklung und Innovation bilden.

Die Abbildung 11 beinhaltet ein Beispiel für nationale FuE-Daten der Internet-Veröffentlichung, Abbildung 12 die grafische Umsetzung internationaler Vergleichszahlen.

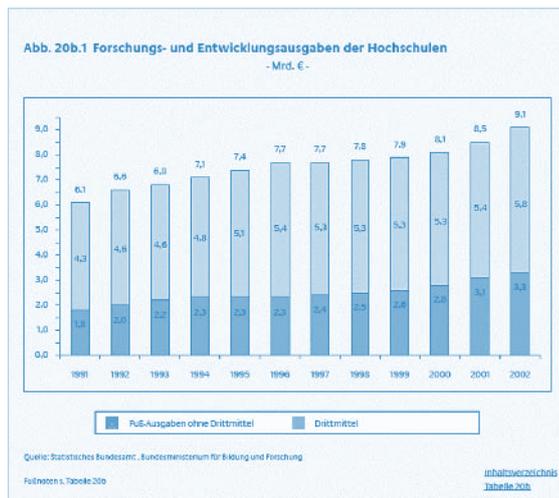
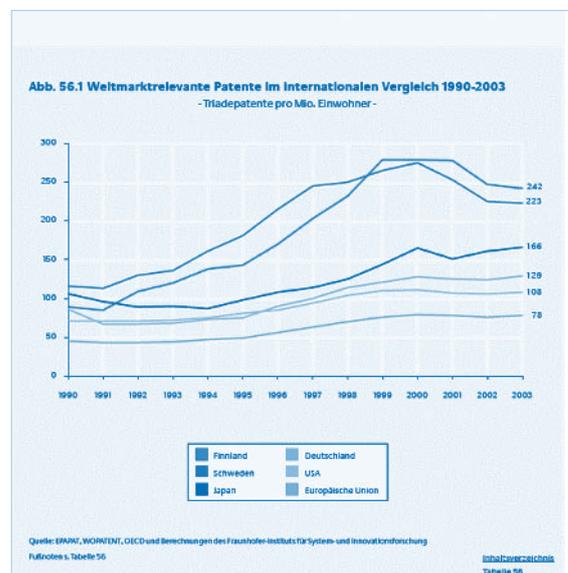


Abb. 11: Grafik aus der Internet-Veröffentlichung des BMBF (national)

Abb. 12: Grafik aus der Internet-Veröffentlichung des BMBF (internationaler Vergleich)



3.2 Datenlogistik

Die FOSTAT-Daten werden aus den unterschiedlichsten Quellen bezogen, die meisten davon werden im BMBF zentral abgerufen und an die HIS zur Weiterverarbeitung in der Datenbank übermittelt, einige internationale Vergleichsdaten werden direkt per Internet beispielsweise von Eurostat abgefragt. Die Hauptquellen sind

- Das BMBF selber, z.B. mit Hilfe von DASTAT
- Das Statistische Bundesamt, u.a. für regionale und hochschulbezogene Daten, FuE-Personal sowie Bildungsindikatoren
- Die Stifterverband Wissenschaftsstatistik GmbH für Daten der Wirtschaft
- Das Zentrum für Europäische Wirtschaftsforschung (ZEW) und das Niedersächsische Institut für Wirtschaftsforschung (NIW) für Unternehmens- und Wirtschaftsentwicklungen
- Internationale Quellen (EU, OECD) für staatenübergreifende Vergleichsdaten
- Die Deutsche Bundesbank (Zahlungsverkehr)

Alle FOSTAT-Tabellen bilden jahresbezogene Zeitreihen ab und werden für jede anstehende Veröffentlichung aktualisiert. Ähnlich wie DASTAT ist auch die FOSTAT-Datenbank in stichtagsbezogenen Ständen organisiert, wobei ein Stand jeweils einer Veröffentlichung entspricht. Die Texte und Beschriftungen liegen in deutscher und englischer Sprache vor.

Für einen neuen Stand werden die FOSTAT-Daten teilweise nur um aktuellere Werte ergänzt, andere dagegen werden komplett neu importiert. Der Bedarf, Daten vollständig durch andere zu ersetzen, tritt insbesondere auch dann auf, wenn sich generelle Zuordnungen oder Indikatoren ändern (etwa die Revision der Volkswirtschaftlichen Gesamtrechnung im April 2005 [9] oder Änderungen am Zuschnitt von Bundesressorts).

Technisch geschieht das Anlegen und Befüllen eines neuen Standes in der Weise, dass mit Hilfe eines Werkzeugs ein bestehender Stand (oder eine Auswahl von Tabellen eines Standes) zunächst in einen neuen kopiert wird, um dann mit Hilfe halbautomatisierter Mechanismen die neuen Daten zu ergänzen bzw. die alten Daten durch neue zu ersetzen. Eine ausführliche anschließende Sichtung der Zahlen und Beschriftungen nebst Plausibilisierungsmaßnahmen schließt das Verfahren ab und der neue Stand wird für Auswertungen freigegeben.

3.3 Datensegmente

Die FOSTAT-Datenbasis wird derzeit ausschließlich für das Erzeugen vordefinierter Tabellen und teilweise vordefinierter Abbildungen benutzt. Die Datenbank beinhaltet zu diesem Zweck neben den Daten sehr ausführliche Layout-Informationen. Dies gilt insbesondere für die FOSTAT-Tabellen, teilweise aber auch für die Abbildungen (soweit die Eigenschaften der Graphiken fixiert sind, wie etwa der Schriftfont und der Abbildungstyp).

In der Entwicklung befindet sich derzeit eine komplexe Neuausrichtung des Systems zu einem Data Warehouse, bestehend aus einer Datenbasis, welche alle verfügbaren Daten beinhaltet (also nicht nur die zu publizierenden) und die für flexible Ad-hoc-Anfragen zur Verfügung stehen soll. Aus der Datenbasis werden periodisch die jeweils in Form von Tabellen und Abbildungen zu veröffentlichen Einheiten (für ausgewählte Jahre) entnommen und, mit Layout-Informationen versehen, in separaten Segmenten der Datenbank abgespeichert (Abbildung 13).

Wie die vordefinierten Standardtabellen können auch die Ergebnisse der Ad-hoc-Anfragen in unterschiedlichen Formaten abgerufen werden (darunter Microsoft-Excel), um sie anschließend weiter verarbeiten zu können. Dieses Verfahren hat sich nicht nur bei DASTAT bewährt, sondern auch im HIS-System ICE ([10]).

Eine Zwischenstufe zwischen völlig flexiblen Anfragen an die Datenbasis und dem Abruf standardisierter Tabellen bildet die jahresweise Einschränkung von Anfragen. Bei diesem Verfahren werden die Tabellen nicht vollkommen determiniert aus dem betreffenden Datensegment für Publikationen abgerufen, sondern für flexible Zeiträume. Von der Quantität her bietet auch diese Stufe einen verbreiterten Zugriff auf die Daten, denn publiziert wird jeweils lediglich ein kleiner zeitlicher Ausschnitt aus den verfügbaren Zahlen.

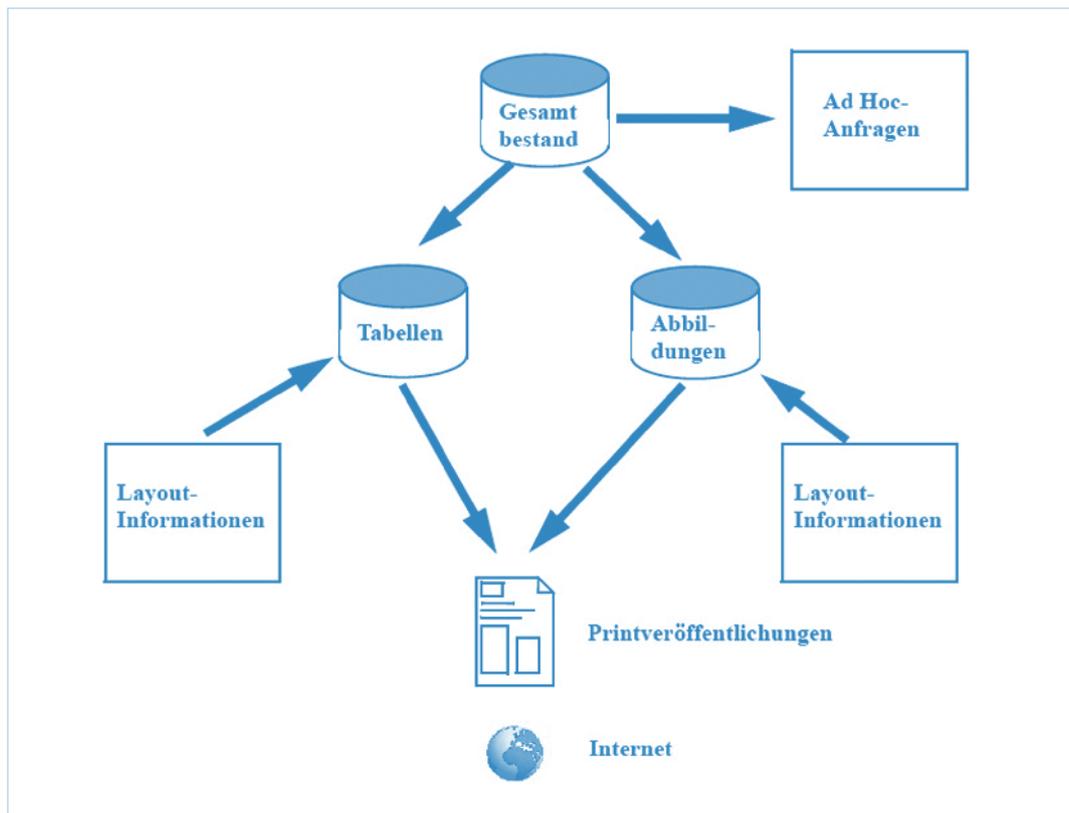


Abb. 13: FOSTAT-Datensegmente

3.4 Systemarchitektur und Anwendungskomponenten

FOSTAT-Tabellen besitzen teilweise sehr komplexe Layouts, etwa horizontale und/oder vertikale Zentrierungen über mehrere Spalten, sehr spezielle Schriftvorgaben, festgelegte Hintergrundfarben, usw. Einen Ausschnitt aus einer Beispieltabelle für eine Internetveröffentlichung zeigt Abbildung die 14. All diese Informationen sind neben den Tabellen und Beschriftungen in der Datenbank gespeichert und werden beim Abrufen einer Tabelle per JDBC (Java Database Connectivity) ausgelesen und weiterverarbeitet, wobei zahlreiche Layout-Informationen für jedes in FOSTAT

mögliche Ausgabeformat getrennt gespeichert sind (Abstandsdefinitionen sehen z.B. in HTML anders aus als in PDF).

Jahr	finanziert durch				FuE-Ausgaben insgesamt ²⁾	
	Gebietskorperschaften ^{3) 3)}		Wirtschaft ⁴⁾	Private Institutionen ohne Erwerbszweck ⁵⁾	Mio. €	in % des BNE ⁷⁾
	Mio. €	in % des öffentlichen Gesamthaushalts ⁶⁾	Mio. €			
1981	8 981	3,2	11 154	78	20 214	2,5
1983	9 475	3,2	13 011	86	22 571	2,6
1985	10 587	3,4	15 896	68	26 551	2,8
1987	11 114	3,3	18 831	122	30 067	2,9
1989	11 864	3,3	21 064	166	33 094	2,8
1991	14 821	3,2	23 935	196	38 952	2,5
1993	15 492	2,7	23 973	122	39 567	2,3
1995	15 735	2,6	24 733	104	40 572	2,2
1997	15 608	2,6	27 036	141	42 785	2,2
1999	15 956	2,6	32 411	205	48 572	2,4
2001	16 820	2,8	35 095	222	52 137	2,5
2003	17 128	2,8	38 060	176	55 364	2,6

Abb. 14 FOSTAT-Tabelle für eine Internetveröffentlichung (Auszug)

Der Benutzer sieht beim Aufruf der FOSTAT-Auswertungen zunächst ein Java-Applett, in dem er Anmeldeinformationen eingibt und einen Stand für die gewünschten Tabellenabfragen vorgibt. Anschließend kann eine Tabellenummer und ein Ausgabeformat ausgewählt werden. Diese Informationen werden an den Server übermittelt, der die angefragte Tabelle dynamisch erzeugt und an den Benutzer zurück als Antwort schickt.

Serverseitig wird das Cocoon XML Publishing Framework der Apache Group ([11]) eingesetzt, wobei die Daten und Layout-Informationen mit Hilfe einer FOSTAT-spezifischen Geschäftslogik in aufbereiteter Form an den Cocoon-Prozessor übergeben und mit zusätzlichen, in einem speziellen XSL-Stylesheet abgelegten statischen Informationen versehen werden, um daraus eine XML-Datei zu erzeugen, die schließlich durch Transformatoren in das gewünschte Format (z.B. PDF für den Adobe Acrobat Reader oder XLS für das Microsoft Excel-Programm) überführt wird (vgl. Abbildung 15).

Hinsichtlich der Architektur hat man es somit auch hier mit einem Dreischichtensystem zu tun (Datenbank, Applikationsserver, Client), wobei die Datenbank physisch als Kopie auf mehreren Rechnern liegt und sogar mit mehreren Instanzen auf einzelnen Servern, die jeweils durch kleine Konfigurationsänderungen angesprochen werden können. Auch der Applikationsserver liegt in replizierter Form mehrfach vor, da man für Entwicklungen immer eine andere Umgebung benutzen wird als für den produktiven Einsatz. Alle Server befinden sich in den Räumen der HIS GmbH, wobei jeweils Linux-Betriebssysteme auf den Servern laufen und PostgreSQL als relationales Datenbankmanagementsystem.

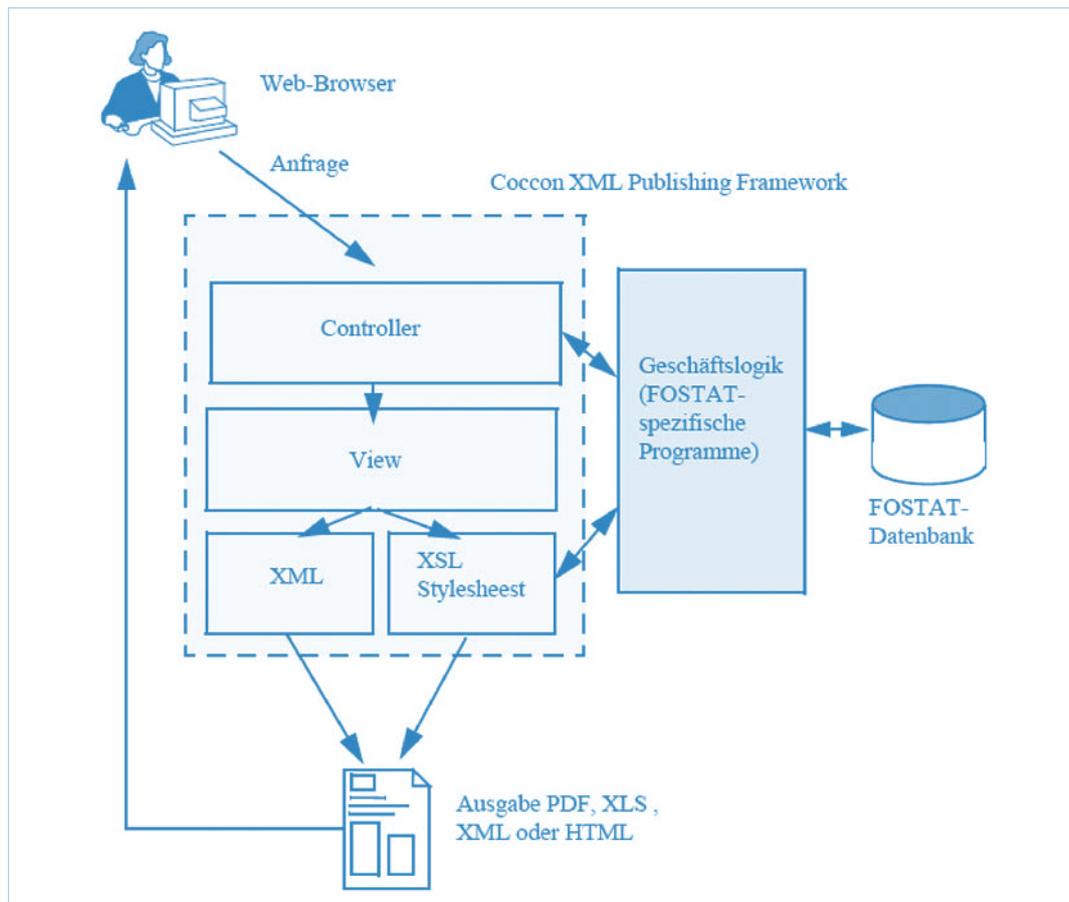


Abb. 15: FOSTAT im Kontext des Cocoon XML Publishing Framework

FOSTAT-Abbildungen können aus einem vorhandenen Pool vordefinierter Graphiken abgerufen werden (derzeit in Absprache mit dem Auftraggeber BMBF 30 Abbildungen, für die Zukunft sind etwa 100 Abbildungen vorgesehen). Auch die Abbildungen werden mit Hilfe des Cocoon-Frameworks, hier zusätzlich unter Einbeziehung der Open Source-Klassenbibliothek JFreeChart erzeugt ([12]). Eine große technische Herausforderung bei derartigen Klassenbibliotheken besteht in der Tatsache, dass Abbildungen im Normalfall auf dem Rechner des Benutzers generiert werden, da sie dort auch angezeigt werden. In FOSTAT dagegen sollen die Abbildungen auf dem Server erzeugt und anschließend direkt nach PDF konvertiert und in diesem Format zum Benutzer übermittelt werden. Es waren daher viele und teilweise komplexe Anpassungen und Eigenentwicklungen nötig, um diese Form der Graphikverarbeitung zu erreichen.

Die Abbildungen 16 und 17 zeigen zwei Beispiele für FOSTAT-Anwendungswerkzeuge.



Abb. 16: FOSTAT-Standverwaltung

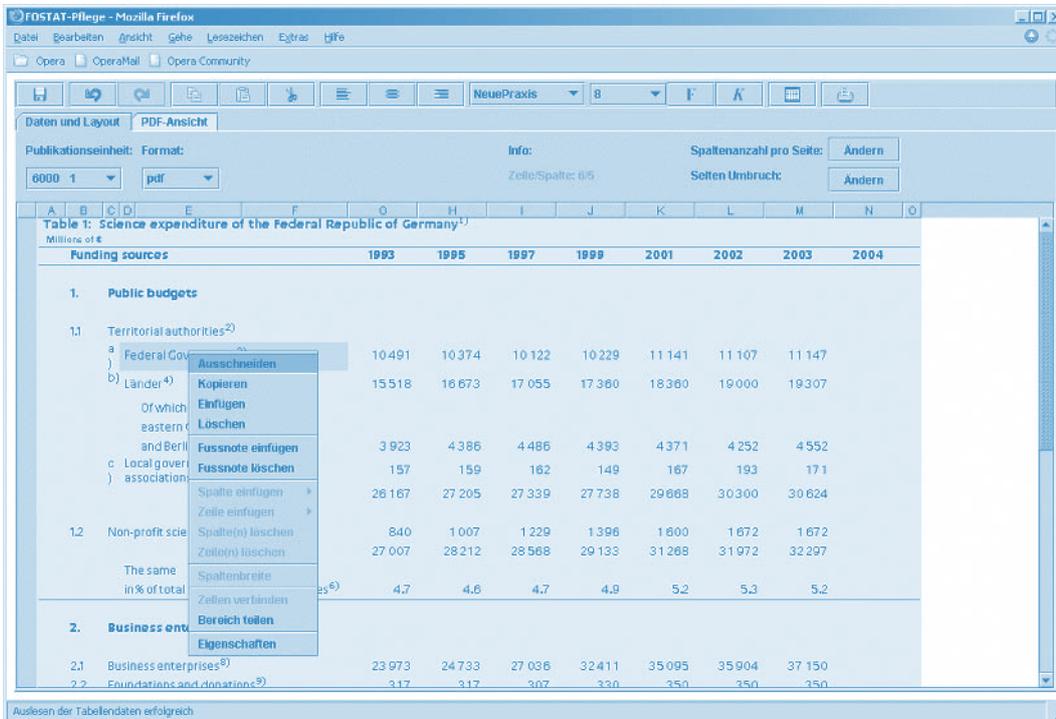


Abb. 16: FOSTAT-Datenpflege-Werkzeug

4. Ausblick

DASTAT ist seit mehreren Jahren im produktiven Einsatz und die Entwicklungen sind funktional aus derzeitiger Sicht nahezu vollständig (bei aller Problematik einer solchen Aussage bezogen auf Software). Auf der anderen Seite äußern die Anwender des öfteren Wünsche nach zusätzlichen Komponenten oder Teilkomponenten, bzw. nach geänderter Handhabung oder der Einbeziehung neuer Datenbereiche. Neben der Umsetzung solcher Wünsche besteht eine Anforderung an die Entwicklergruppe darin, die Anwender inhaltlich zu unterstützen (etwa bei besonderen Einzelanfragen direkt an die Datenbank, für welche die DASTAT-Software nicht konzipiert ist, oder bei organisatorischen Problemen im Kontext der Ressorthebungen) und das Gesamtsystem technisch immer auf einer aktuellen Grundlage zu erhalten. Daneben steht die Projektgruppe Anwenderschulungen in den Ministerien laufend zur Verfügung.

FOSTAT dagegen wird derzeit auf mehreren Feldern weiterentwickelt, insbesondere in Bezug auf die flexiblen Abfragemöglichkeiten des Data Warehouse. Auch die Pflege- und Verwaltungstools werden laufend verbessert. Ein weiteres Aktionsfeld der Projektgruppe besteht in der Zusammenführung der Daten zu einer Gesamtpublikation, bzw. der Konvertierung von FOSTAT-Ausgaben in Formate, die von der jeweiligen Agentur für die Printveröffentlichungen weiter verarbeitet werden können (z.B. spezielle EPS-Formate und druckerbezogene Farbkodierungen).

Für die Zukunft ist der Einsatz der hier vorgestellten Data Warehouses oder einiger Komponenten auch für weitere Nutzerkreise denkbar, z.B. auf der Länderebene.

5. Quellen

- [1] <http://www.euractiv.com/Article?tcmuri=tcm:31-103671-16&type=LinksDossier&textlg=DE>
- [2] **Forschung und Innovation in Deutschland 2006** http://www.bmbf.de/pub/forschung_und_innovation_in_deutschland_2006.pdf
- [3] <http://www.kp.dlr.de/profi/easy/formular.html>
- [4] **Main Definitions and Conventions for the Measurement of Research and Experimental Development (R&D), a Summary of the Frascati Manual**, OECD Paris 1994
- [5] **Oslo Manual: The Measurement of Scientific and Technological Activities**, OECD 1992
- [6] H. Dicken: **JDBC: Internet-Datenbank-Anbindung mit Java**, International Thomson Publishing, Bonn, 1997
- [7] H. Dicken, D. Huke, R. Schmelzer: **Information at your fingerprint: Informationssysteme für die Forschungsstatistik**, Javasppektrum, 01/05, SIGS Conferences GmbH, München 2005

- [8] Bundesministerium für Bildung und Forschung: **Bundesbericht Forschung**, Bonn/Berlin 2004

- [9] <http://www.destatis.de/presse/deutsch/pm2005/p1970121.htm>

- [10] P. Müßig-Trapp, H. Dicken, H. Kopp: **ICE - a Web Based Information System for Decision Support in Higher Education Politics**, 9th International Conference of European University Information Systems, Amsterdam, 2-4 July 2003

- [11] <http://cocoon.apache.org>

- [12] <http://www.jfree.org/jfreechart/>

HIS, Goseriende 9, 30159 Hannover

Postvertriebsstück, Deutsche Post AG, Entgelt bezahlt, 61246

Herausgeber:

HIS-Hochschul-Informationen-System GmbH
Goseriede 9, 30159 Hannover
www.his.de

Verantwortlich:

Prof. Dr. Martin Leitner

Hinweis gemäß § 33 Datenschutzgesetz (BDSG):

Die für den Versand erforderlichen Daten (Name, Anschrift) werden elektronisch gespeichert.

ISSN 1863-5563

