

2 Datenbank - Grundlagen

Dieses Kapitel vermittelt einen Überblick über die Begriffswelt der Datenbanken. Die Begriffe werden in den vertiefenden späteren Kapiteln verwendet.

2.1 Betrachtungsebenen für Informationen und Daten, Datenschema

Der Begriffe Information und Daten wurden bereits angesprochen. In der Datentechnik können Daten auf fünf zugeordneten Ebenen betrachtet werden. (Bild2.1).

Für die Anwender sind die Ebenen 1 und 2 wichtig (in Bild2.1 zuoberst). Die Ebene 1 ist die Ebene der *realen Welt*.

Das Beispiel:

Heidi Meier leiht sich ein rares Exemplar einer Bibel aus. „Heidi Meier“ und „diese Bibel“ gehören der realen Welt an.

Auf der Ebene 2, der sog. *Informationsebene*, sollen bestimmte Fragen über die reale Welt beantwortet werden können. Dazu muss die reale Welt durch Daten beschrieben und modelliert werden. Zur Umschreibung der realen Welt werden Begriffsbezeichnungen und -inhalte (Wörter, Zahlen, Bilder, Töne usw.) verwendet. Mit Computertechnik hat dies noch nichts zu tun.

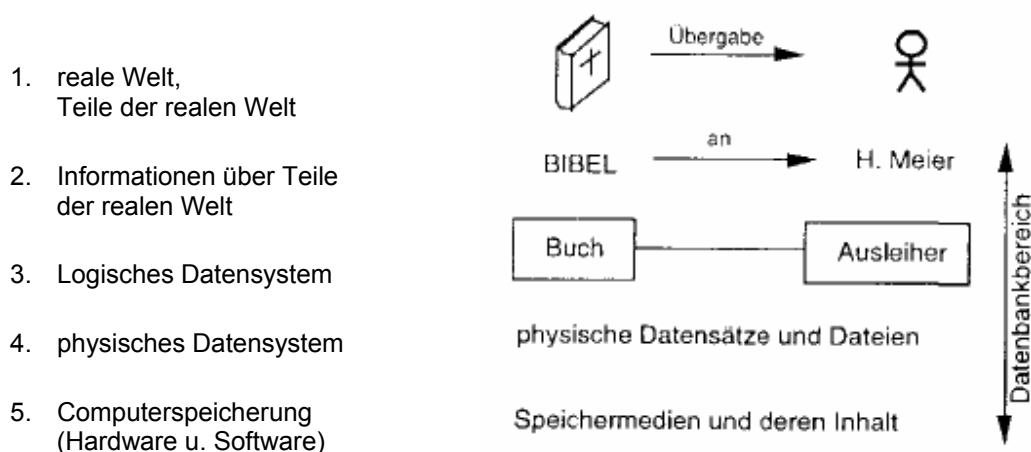


Bild 2.1 Betrachtungsebenen für Daten am Beispiel „Buchausleihe“

Die Technik steht auf den Ebenen 4 und 5 im Vordergrund.

Die Abstraktion der Ebene 3 ermöglicht den Schritt zur Ebene 4 der *physischen Dateiorganisation*.

Auf der Ebene 5 werden die Daten *computertechnisch* betrachtet. Hier geht es *etwa* um Plattenspeicher und deren technische Eigenschaften wie Speicherdichten und -anordnungen.

Beide untersten Ebenen sind das ausschliessliche Arbeitsfeld des systemnahen Informatikers und interessieren den Anwender kaum.

Anwender und Informatiker treffen sich in der Mitte, also auf der logischen *Ebene*. Auf dieser Ebene 3 wird die Realität (z.B., Buch, „Standortnummer“), in einer Form beschrieben, die auch vom Benutzer nach kurzer Einführung verstanden werden kann und die trotzdem so präzise ist, wie dies der Informatiker benötigt. Diese Beschreibung erstellen Anwendervertreter gemeinsam mit Informatikern. Neu zu erstellende und zu ändernde Datenbanken werden auf dieser logischen Ebene definiert.

Das Ergebnis dieser Entwurfsarbeit ist das sog. Datenschema:

Definition.:

Ein ***Datenschema*** oder kurz **Schema** ist eine Beschreibung eines Datenbestandes. Das Datenschema beinhaltet die Metadaten zum Datenbestand. Es **definiert den künftigen Inhalt der Datenbank, die Nutzdaten**.

Definition.:

Nutzdaten ist die Bezeichnung für die Gesamtheit der gespeicherten Datenwerte, die den Inhalt der Datenbank bilden.

2.2 Strukturierte, konsistente und redundante Daten

In einer Datenbank werden typischerweise viele gleichartige *Datensätze* zu einem bestimmten Aufgabenbereich abgespeichert, also etwa Adressen zu einem Adressverzeichnis, Kunden zu einer Kundenliste, Flüge zu einem Flugplan. Jeder *Datensatz* bezieht sich auf *ein* derartiges Element (auch „Entität“ genannt; die genaue Definition des Begriffs „Entität“ folgt noch.).

Beispiele: Eine Adresse, ein Kunde, ein Flug.

In Bild 2 entspricht jede Zeile der gezeigten Tabellen einem Datensatz, im einen Fall einer 'Adresse', im andern Fall einer 'Poststelle'.

Name	Vorname	Strasse+Nr	PLZ	Ortschaft
Meier	Heidi	Badstrasse 3	5400	Baden
Müller	Patrick	Buckmatte 1	5400	Münzlishausen
Roth	Ursula	Zwyssigstrasse 28	5400	Wettingen

PLZ	Ortschaft
5400	Baden
5400	Münzlishausen
5405	Baden-Dättwil
5430	Wettingen

Bild 2 Beispiel für strukturierte Daten

Die einzelnen *Datensätze* bestehen je aus mehreren *Datenfeldern* und beschreiben so wichtige Eigenschaften der entsprechenden Entität.

Beispiel:

Ein *Adressdatensatz* enthält Datenfelder für Name, Vorname, Anrede, Straße und Hausnummer, Postleitzahl, Ortsname.

Das einzelne *Datenfeld* wiederum enthält Datenwerte eines bestimmten *Wertebereichs*.

Beispiel:

Die Postleitzahl ist in Deutschland eine fünfstellige ganze Zahl, in Österreich und der Schweiz eine vierstellige ganze Zahl.

Datenbestände können aber nicht nur immer feiner aufgegliedert werden (Tabelle -> Datensatz -> Datenfeld), sondern sie können zusammengesetzt und miteinander verknüpft werden.

Beispiel:

In Bild 2 werden zwei Tabellen indirekt verknüpft, indem die Postleitzahlen (PLZ) in der Tabelle „Adressen“ nur Werte annehmen dürfen, die als PLZ in der Tabelle „Poststellen“ vorkommen. (In Bild 2 steckt übrigens ein Fehler; haben Sie ihn erkannt?).

All diese Aufgliederungen und Zusammensetzungen sind typisch für viele Datenbestände, die auf Computern bearbeitet werden, ganz besonders für Daten in Datenbanken. Solche Datenbestände werden *strukturiert* genannt

Definition:

Strukturiert heißt ein Datenbestand, bei dem systematische Untergliederungen und Verknüpfungen möglich sind.

Längst nicht alle Datenbestände sind übrigens strukturiert. Ein Gegenbeispiel bietet etwa der Text eines Kriminalromans. Der ganze Roman besteht im wesentlichen aus einer einzigen Sequenz von Schriftzeichen, nicht aber „aus vielen gleichartigen Datensätzen“. In einem Roman liest man keine isolierten Textabsätze einzeln.

Die Daten einer Datenbank sind nicht nur strukturiert, sondern sie müssen meist zusätzlich bestimmten Bedingungen, den sogenannten *Konsistenzbedingungen* gehorchen. So müssen die einzelnen Werte in den Datenfeldern bestimmten Wertebereichen entnommen werden (Beispiel: gültige Postleitzahl), bei Verknüpfungen müssen die Werte zusammenpassen und zu sog. *Aussagen* („Heidi hat blaue Augen“) dürfen keine Widersprüche auftreten („Heidi hat braune Augen“). (In Bild.2 enthalten die jeweils untersten Zeilen der beiden Tabellen einen Widerspruch: Hat „Wettingen“ nun die Postleitzahl „5400“ oder „5430“?)

Definition:

Konsistent heißen Daten, wenn sie den für einen Datenbestand vorgegebenen **Konsistenzbedingungen** genügen.

Beispiel:

Die Daten in Bild 2 sind *nicht* konsistent, weil sie widersprüchliche Aussagen enthalten.

Und noch einen weiteren Begriff wollen wir hier erstmals einführen:

Definition:

Redundanz heißt das **mehrfache Vorkommen der gleichen Aussage**.

Offensichtlich stecken in der Poststellentabelle und in der Adresstabelle in Bild 2.2 zum Teil die gleichen Aussagen (Bsp. „Baden hat die Postleitzahl 5400“). Trotzdem wünscht die Post ausdrücklich, daß Adressen Postleitzahl *und* Ortsnamen enthalten, damit bei Schreibfehlern und mehrfachen Ortsbezeichnungen der Adressat besser gefunden werden kann. Umgekehrt können auf Grund der vorhandenen Redundanz jedoch auch Konsistenzfehler auftreten (wie hier in Bild 2.2). Redundanz ist also eine Eigenschaft mit Vor- und Nachteilen.

2.3 Sprachen für die Datenbeschreibung

Wer Daten und Datenstrukturen beschreiben will, benötigt dazu ein geeignetes Beschreibungswerkzeug, eine *Datenbeschreibungssprache* (*data description language* = DDL); sie dient zur *Beschreibung* des Datenschemas.

In der eher theoretischen Datenbankliteratur werden solche Beschreibungssprachen auch *Datenmodelle* genannt; genauer müsste man sagen *informatikbezogene Datenmodelle*. So spricht man etwa vom „Relationenmodell“ oder vom „ER-Modell“ (beide werden in Kap.3 beschrieben).

Der Begriff *Datenmodell* wird gleichzeitig von vielen Datenbankpraktikern in einer zweiten, wesentlich anderen Bedeutung verwendet. Anwender bezeichnen oft die Datenbeschreibung selber, also das *Datenschema*, als *Datenmodell*. Diese Doppelverwendung des Begriffs Datenmodell kann zu Missverständnissen führen, weshalb hier vorzugsweise die eindeutigen Begriffe *Datenbeschreibungssprache* und *Datenschema* verwendet werden.

(Datenschema und Datenmodell werden hier vereinbarungsgemäß mit gleicher Bedeutung verwandt.)

Eine Datenbeschreibungssprache soll wichtige Arten von Datenstrukturen einfach

darstellen können. Zu den wichtigsten *Datenstrukturen* gehören Hierarchien, Netzwerke und Relationen, die nachstehend kurz vorgestellt werden. Dabei sind Strukturen nicht Selbstzweck, sondern dienen dazu, reale Datensysteme darzustellen, zu modellieren. Modellbildung ist gleichbedeutend mit dem Herausarbeiten von Strukturen.

Als *Beispiel* betrachten wir den Bücherkatalog einer kleinen Bibliothek. Dabei wird jedes Buch durch seine Katalogkarte repräsentiert, und jedes Buch gehört zu einem Fachgebiet. Diese Fachgebietsbeziehung soll nun in den drei verschiedenen Datenstrukturen dargestellt werden.

In den folgenden Darstellungen müssen immer die konkreten Datenbestände („Vorkommen“ oder „Instanzen“, links) und ihre Struktur (Schema, rechts) auseinandergehalten werden.

Hierarchien

Hierarchische Gliederungen von Datenbeständen werden in der Datenverarbeitung vielfach verwendet.

Bild 2.3 zeigt links ein Beispiel einer *Hierarchie*:

Zu jedem Fachgebiet gehören mehrere Bücher, zu jedem Buch genau ein Fachgebiet.

Die Strukturdarstellung (Bild 2.3 rechts) nennt eine Hierarchie eine 1 - m-Beziehung (sprich „1 zu m“). Diese sagt aus, dass zu jedem Fachgebiet mehrere (m) Bücher („Kinder“) gehören, zu jedem Buch aber genauem (1) Fachgebiet („Vater“).

Hierarchische Strukturen (auch mehrstufige) sind einfach zu überblicken und erlauben eine sehr rationelle und klare Organisation der Daten, nämlich eine sequentielle Speicherung und Verarbeitung.

ng.

Dabei erfolgt die Ordnung nach dem Satz „Alle Söhne nach dem Vater“. Diese „natürliche Reihenfolge“ kommt später nochmals zur Sprache.

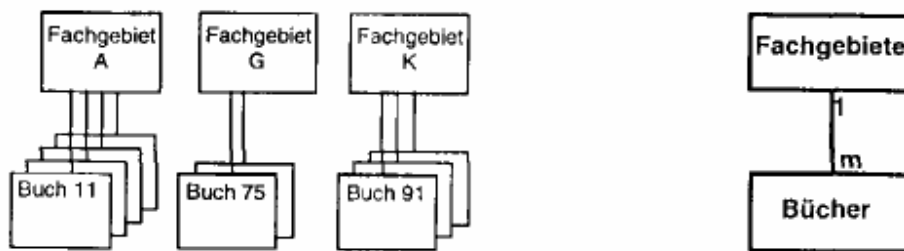


Bild 2.3 Hierarchie: Elemente- und Strukturdarstellung

Hierarchische Strukturen sind sehr verbreitet, aber in ihrer Modellierungsfähigkeit recht beschränkt. Namentlich erlauben sie nicht, Netzwerke direkt darzustellen.

Netzwerke

In unserem Beispiel Bücherkatalog kann es sehr wohl vorkommen, dass ein bestimmtes Buch gleichzeitig zu mehreren Fachgebieten gehört. Dies lässt sich mittels einer Hierarchie (Bild 2.3) nicht ausdrücken, sondern erfordert ein Netzwerk (Bild 2.4), eine sog. m - m Beziehung (gelesen als „m-zu-m-Beziehung“). Das Symbol „m“ steht dabei nicht für eine bestimmte Zahl, sondern sagt nur, dass es „mehrere (many) Elemente“ sein können. In einer m - m-Beziehung müssen die beiden „m“ nicht für den gleichen Wert stehen. Die Strukturdarstellung von Bild 2.4 gibt somit an, dass zu jedem Fachgebiet mehrere Bücher, gleichzeitig aber auch zu jedem Buch mehrere Fachgebiete gehören können.

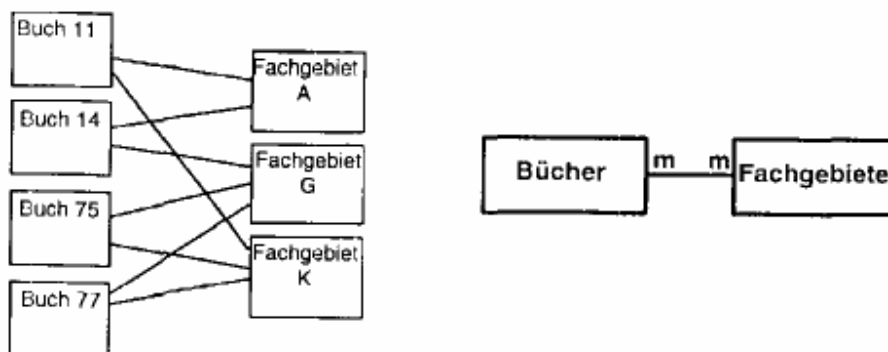


Bild 2.4 Netzwerk: Element- und Strukturdarstellung

Jede m - m-Beziehung lässt sich durch *zwei* Hierarchien *zusammen* darstellen (wie später noch detailliert gezeigt wird). Dabei werden die einzelnen Beziehungen (die Verbindungslinien in Bild 2.4 links) als eigenständige Elemente aufgefasst und je in einer 1 - m Beziehung den Büchern und den Fachgebieten untergeordnet.

Durch diese gleichzeitige Einordnung der Elemente in zwei *verschiedene* Hierarchien lassen sich diese allerdings nicht mehr sequentiell eindeutig einordnen, so dass wir schon an dieser Stelle festhalten können, dass unter den Elementen eines Netzwerks keine natürliche Reihenfolge existiert. Daher lassen sie sich nicht so effizient wie Hierarchien verarbeiten und auch nicht so einfach darstellen.

Tabellen (Relationen)

Tabellen bilden seit jeher (also schon lange vor dem Zeitalter des *Computers*) eine wichtige Darstellungsform für Daten. Dabei kommen Tabellen nicht nur in der zweidimensionalen Grundform (Bild 2.5) vor, sondern auch in komplizierteren Formen, denen wir später noch begegnen werden. Eine besonders wichtige Form ist aber die Grundform der zweidimensionalen Tabelle mit je einem Datenwert pro Tabellenfeld (Bild 2.5); jede Kolonne (Spalte) entspricht einem darzustellenden *Merkmal* oder *Attribut* (Bsp.: Autor, Titel), jede Zeile einer darzustellenden Einheit (Bsp.: einem *Buch*). Diese Darstellung wurde schon in Bild 2.2 verwendet.

Zeile = Datensatz (*Tupel*).

Die Datenwerte eines Datensatzes beziehen sich auf die gleiche Einheit der Realität; sie gehören somit logisch zusammen. Darüber hinaus lassen sich in einer einzelnen Tabelle keine Beziehungen explizit darstellen.

Beziehungen lassen sich jedoch indirekt darstellen, indem die *gleichen* Merkmale in *mehreren* Tabellen *vorkommen* können. So muss in unserem Beispiel neben der Tabelle „Bücher“ (Bild 2.5) eine zweite Tabelle „Fachgebiete“ mit den Merkmalen Fachgebietsbezeichnung und Fachgebietsnummer [FG-Nr.] aufgebaut werden; die Bedeutung der Fachgebietsnummer 5 = „Schauspiel“ muss beim Aufruf in einer separaten Tabelle „Fachgebiet“ nachgesehen werden. Eine solche Lösung mit mehreren Tabellen bedeutet einfachere Speicherung, aber Mehraufwand bei Abfragen, denn Beziehungen müssen jeweils durch systematischen Vergleich von Werten in verschiedenen Tabellen rekonstruiert werden.

Bücher

Buch-Nr	FG-Nr.	Autor:	Titel:	Jahr:
11	S	Dürrenmatt F.	Die Physiker	1962
14	S	Goethe J.W.	Faust I	1806
75	R	Kästner E.	Fabian	1931
98	S	Schiller F.	Wilhelm Tell	1802

Datensatz

Attribut

Bild 2.5 Eine Tabelle (oder Relation)

Die Darstellung von Daten in Tabellen bildet die Grundlage des sog. *Relationenmodells*, das noch ausführlich zur Behandlung kommt.

Das Relationenmodell geht auf E.F. Codd zurück, der seine Arbeiten dazu ab 1970 publizierte [Codd 70]. Es hat seine Bedeutung wegen der Systematik der Datendarstellung (mathematische Klarheit) und der Entkopplung der Datenbereitstellung von ihrer zukünftigen Verwendung. Hingegen war das Relationenmodell in ursprünglicher Form nur beschränkt praktisch einsetzbar, vor allem aus Effizienzgründen. Dies hat sich aber inzwischen geändert; moderne Datenbanken in Großanwendungen basieren meist auf dem Relationenmodell. Die Tabelle als Datenstruktur ist aber schon viel älter als das Relationenmodell; sie hat jedoch mit dem Relationenmodell eine neue Bedeutung erlangt.

Das Relationenmodell wird im Rahmen des Datenschema-Entwurfs (Kap. 3) ausführlich behandelt.

2.4 Sprachen für die Datenmanipulation

Für die Arbeit mit Datenbanken braucht es nicht bloß eine systematische Beschreibung der Daten und ihrer Strukturen, sondern auch eine Beschreibung der Arbeitsmöglichkeiten mit diesen Daten:

Speichern, Suchen und Lesen. Datenmanipulation heißt der Sammelbegriff für all diese Tätigkeiten.

Datenpflege (mutieren): einfügen, ändern, entfernen

Bei der Datenpflege werden in einer Datenbank einzelne Datensätze neu eingefügt,

geändert oder entfernt. Am Beispiel „Bücherkatalog“ in Abschnitt 2.3 sieht das etwa so aus:

Einfügen: Ein neues Buch (bzw. der Datensatz, der diesem Buch entspricht) wird beschrieben (Autor, Titel, Fachgebiet(e), ...) und (automatisch) in die entsprechende Datenstruktur eingefügt, und zwar

in einer Hierarchie (Bild 2.3) als Element unter dem entsprechenden Fachgebiet,

in einem Netzwerk (Bild 2.4) als Element samt den Beziehungen zu den entsprechenden Fachgebieten,

in einer Tabelle (Bild 2.5) als zusätzliche Zeile (auf die Darstellung von Beziehungen bei Tabellen wird erst in Kap. 3 vertieft eingegangen).

Ändern: In einem Buchdatensatz wurde ein Fehler entdeckt („Göthe“ statt „Goethe“) und soll behoben werden. Dazu wird der betroffene Datensatz geholt, das falsche Datenfeld geändert und der korrigierte Datensatz an der bisherigen Speicherstelle wieder gespeichert. Der Speicherbedarf ändert sich nicht; eine Änderung verläuft somit für Hierarchien, Netzwerke und Tabellen gleichartig (sofern nicht Beziehungen von der Änderung mitbetroffen sind).

Entfernen, löschen: Ein defektes Buch wird weggeworfen, der zugehörige Datensatz entfernt. Bei diesem Löschprozess muss allerdings darauf geachtet werden, dass existierende Beziehungen miteinbezogen werden.

Alle erwähnten Operationen der Datenpflege — einfügen, ändern, entfernen — sind relativ einfach zu verstehen, aber technisch anspruchsvoll. Denn ihre korrekte Abwicklung stellt an jede Datenbank hohe Anforderungen, weil bei diesen Operationen in den Datenbestand eingegriffen wird; er wird verändert.

Datensuche (abfragen, lesen): auswählen, navigieren, darstellen

Bei der Datensuche wird der Datenbestand nicht verändert, sondern bloß nach den Bedürfnissen des Anwenders abgesucht; die Sicherheitsprobleme sind somit bei dieser Operation bedeutend geringer. Andererseits können Suchoperationen bei großen Datenbeständen sehr aufwendig sein.

Auch bei der Datensuche lassen sich verschiedene Grundoperationen unterscheiden.

Auch sie sollen an den Beispielen des „Bücherkatalogs“ (**Abschnitt 2.3**) verdeutlicht werden.

Auswählen: Wer Merkmale des gesuchten Datensatzes (der gesuchten Datensätze) bereits kennt, kann diese formulieren und zur direkten Suche benützen.

Beispiele:

Buch mit Buchnummer = „14“?

Bücher vom Autor „Goethe“?

Bücher vom Autor „Goethe“ und mit Fachgebiet „S“

Navigieren: Bei diesem Vorgehen wird keine vollständig neue Frage gestellt, sondern von einem früher erhaltenen Suchergebnis aus weitergesucht. So hat etwa jemand in einem Bücherkatalog etwas Interessantes gefunden; das Buch ist aber bereits ausgeliehen. Eine nächste Suchfrage könnte daher wie folgt lauten:

Andere Bücher zum gleichen Sachgebiet?

Ein sehr wichtiges Beispiel zum Navigieren ist heute das Suchen im World Wide Web; aber auch in Datenbanken kann navigiert werden.

Das Ergebnis einer Datensuche besteht typischerweise aus einem oder mehreren Datensätzen. Deren Inhalt muss in vielen Fällen noch umgeformt werden, bevor er für den Anwender gut lesbar auf Bildschirm oder Papiausdruck erscheint. Zu den Grundoperationen beim Datenabfragen gehört daher auch das *Darstellen* von Daten.

Sprachen

Damit die Anwender die soeben vorgestellten Operationen benützen können, bietet jede Datenbank eine Reihe von Sprachelementen an, die zusammen eine sog. *Datenmanipulationssprache* (*data manipulation language* = *DML*) bilden. In solchen Datenmanipulationssprachen treten typischerweise die erwähnten Stichworte auf, meist auf englisch (insert, modify, update, delete, select).

Beim Prozess des Auswählens geht es grundsätzlich um die Bestimmung einer Auswahl (=Untermenge) aus einer großen Zahl von Datensätzen (Grundmenge). Es liegt daher nahe, dass viele DML, namentlich die sog. relationalen DML, für die Beschreibung dieser Auswahl das Instrumentarium der Mengenlehre verwenden, also etwa Bedingungen mit Vergleichoperatoren (<, >, =) sowie Mengenoperatoren, wie Vereinigung und Durchschnitt.

Die Datenmanipulation wird noch ausführlich behandelt.

2.5 Eigenschaften von Datenbanken

Für einen Computer-Anwender, der für sich *allein* einen Computer benützt, steht meist das *Bearbeiten* von Daten — also das Schreiben, Rechnen, Drucken — im Vordergrund. Dass das Computersystem diese Daten — in der Form von Dokumenten und Dateien — auch *speichern* muss, ist für den Anwender selbstverständlich. Der Anwender kümmert sich aber wenig um die Datenverwaltung, da er dabei vom Betriebssystem umfassend unterstützt wird, das namentlich Dokumente und Dateien verwaltet.

Anders bei größeren Informatikanwendungen. Im betrieblichen Alltag und namentlich dort, wo mehrere Personen die gleiche Computeranwendung benützen, bildet meist ein gemeinsamer *Datenbestand* den Kern der ganzen Informatiklösung. Solche Datenbestände (etwa über Produkte, Kunden, Personal) haben meist eine längere Lebensdauer, und verschiedene Anwender benützen sie gemeinsam (Bild 2.6). Diese Datenbestände benötigen daher eine *gemeinsame Datenverwaltung*.

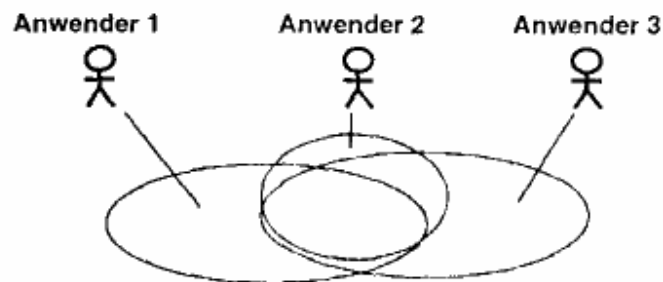


Bild 2.6 Mehrere Anwender mit teilweise gemeinsamen Datenbeständen

Wenn nun ein auf Dauer angelegter Datenbestand einheitlich organisiert, geschützt und verschiedenen Anwendern zugänglich gemacht wird, bilden Daten und Datenverwaltung zusammen eine Datenbank.

Definition:

Eine *Datenbank* ist eine selbständige, auf Dauer und für flexiblen und sicheren

Gebrauch ausgelegte Datenorganisation, die einen Datenbestand (Datenbasis) und die dazugehörige Datenverwaltung (das sogenannte Datenbankverwaltungssystem) umfasst.

Definition:

Das *Datenbankverwaltungssystem* (DBMS = *database management system*) ist ein leistungsfähiges Programm für die flexible Speicherung und Abfrage strukturierter Daten.

Eine Datenbank soll

- **ermöglichen**, dass die Daten effizient und sicher gespeichert und abgerufen werden können,
- **vermeiden**, dass jeder Benutzer sich mit der inneren Organisation des Datenbestandes befassen muss,
- **verhindern**, dass einzelne Benutzer unkontrolliert an die Datenbestände gelangen und damit die Integrität der Daten gefährden können.

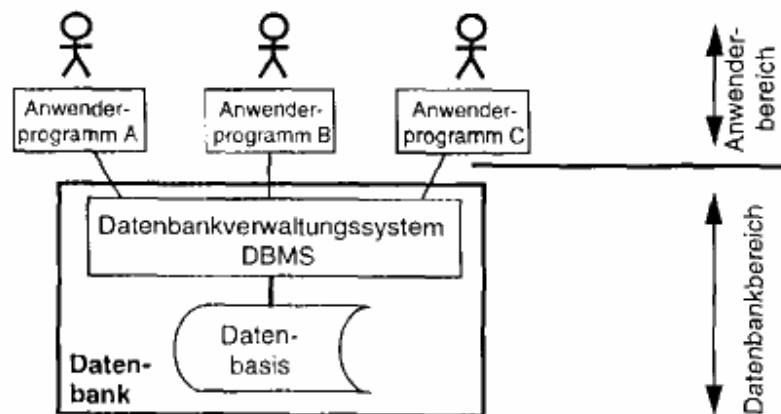


Bild 2.7 Mehrere Anwenderprogramme greifen auf die gleiche Datenbasis zu, und zwar ausschließlich über das DBM – System.

Die Einführung eines zentralen Datenbankverwaltungssystems (DBMS) hat folgende Konsequenzen:

Vorteile: Vereinheitlichung und Zusammenfassung von sonst mehrfach nötigen Funktionen für Datendefinition, Datenorganisation, Datenintegrität; Bereitstellung

effizienter Funktionen für den Zugriff auf Einzeldaten und Datengruppen; einheitliches Datenkonzept; Erleichterung von Systemunterhalt und -weiterentwicklung.

Nachteile: Abhängigkeit von zentralen Konzepten und Entscheidungen; Bereitstellung und Betrieb des Datenverwaltungssystems. (Der *Betrieb* der Datenbank ist allerdings auch *dezentral* möglich; s. später)

Charakteristische Datenbank - Eigenschaften:

Eine Datenbank lässt sich durch folgende Eigenschaften charakterisieren:

A. Strukturierte Daten, keine unnötige Redundanz:

Der Datenbestand besteht aus einer Vielzahl ähnlicher Datensätze mit einem systematischen inneren Aufbau („Struktur“), so dass ein Benutzer sich auf bestimmte Daten und Datengruppen beziehen kann; keine ungeordnete Mehrfachspeicherung (kontrollierte Redundanz).

B. Trennung der Daten von den Anwendungen, was beidseitig unabhängiges Arbeiten ermöglicht:

Datenunabhängigkeit: Die Anwenderprogramme sind „datenunabhängig“; interne Effizienzverbesserungen und Reorganisationen innerhalb der Datenbank tangieren die Anwenderprogramme nicht.

Flexibilität: Die Datenbank dient nicht nur den aktuellen Anwendungen; auch neue Datenbedürfnisse können damit formuliert und befriedigt werden.

C. Datenintegrität:

Die Datenbasis muss hohen Integritätsansprüchen genügen. Dazu dienen systematische Maßnahmen:

- Eingabekontrolle zur Vermeidung der Speicherung widersprüchlicher Daten (Datenkonsistenz).
- Maßnahmen gegen Verlust und Verfälschung der gespeicherten Daten (Verfügbarkeit).
- Maßnahmen gegen unberechtigten Zugriff auf die Daten (Vertraulichkeit).

D. Datenpersistenz:

Die Daten sollen auf Dauer nutzbar sein, auch wenn beteiligte Programme und Computer zeitweise ausgeschaltet werden oder ausfallen.

E. Transaktionsorientiertes Arbeiten:

Die Veränderung von Daten erfolgt ausschließlich mittels korrekter (d.h. konsistenzerhaltender) Transaktionen.

Eine sechste Eigenschaft ist nicht immer, aber oft wichtig:

F. Spezifische Datensicht für verschiedene Benutzer

Der Zugang zur Datenbank lässt sich für jeden Benutzer individuell gemäß seinen Zugangsrechten und -bedürfnissen in Inhalt und Form auf bestimmte Ausschnitte ausrichten.

Nicht alle Datensysteme, die heute als Datenbankverwaltungssystem verkauft werden, verfügen über all die vorstehend genannten Eigenschaften; die Bezeichnung „Datenbank“ kann daher zu falschen Erwartungen führen. Trotzdem ist der Einsatz einfacherer Systeme oft sinnvoll.