

7. Das Data-Dictionary

7.1 Definition und Grundlagen von Data-Dictionaries

Eine Datenbank stellt sich in der Praxis als dynamische Struktur dar. Sie verändert sich häufig kurzfristig im Bereich der gespeicherten Daten und auf lange Sicht auch auf der **Metaebene**, also im Bereich der Beziehungs- und Entity-Typen. Daraus ergibt sich die Notwendigkeit für eine korrekte und vollständige Verwaltung dieser Metaobjekte. Wird dieser Aspekt vernachlässigt, so hat dies in der Regel Effizienzverluste beim Einführen neuer beziehungsweise beim Ändern bestehender Objekttypen zur Folge.

Für die Verwaltung der Metaobjekte kann man einen Systemkatalog, den man gewöhnlich als **Data-Dictionary** bezeichnet, benutzen. Der Begriff **Repository** steht im Rahmen des AD/Cycle-Konzepts der IBM für ein noch umfassenderes Werkzeug, das unter besonderer Berücksichtigung des SAA-Konzeptes weitere Systemdienste zur Verfügung stellt. Das Data-Dictionary ist also unter anderem ein Werkzeug zur Verwaltung der Datenstrukturen. Voraussetzung für die effektive Nutzung eines solchen Systems sind ablauf- und aufbauorganisatorische Maßnahmen. So ist festzulegen, wer in welcher Weise auf die Metadaten zugreifen kann, wer für die Pflege der Metadaten zuständig ist und wie die Daten zur Verfügung gestellt werden. Die Verwaltung eines Data-Dictionaries wird in der Regel von einer zentralen Instanz, der **Datenadministration**, durchgeführt.

Innerhalb eines Data-Dictionaries kann der Datenbegriff sehr weit gefaßt werden. So können neben Records und Feldern auch Programme, Steps, Jobs, Module, Masken und ähnliches zu den Metadaten, die in einem Data-Dictionary verwaltet werden, gehören. Man spricht von **Subjekten**, die ihrerseits zu **Subjektkategorien** zusammengefaßt werden. Ein Data-Dictionary kann alle Subjekte enthalten, die für die Entwicklung von Informationssystemen von Bedeutung sind.

Das Data-Dictionary speichert alle wichtigen Eigenschaften der Subjekte und bietet dadurch eine vollständige Beschreibung. Insbesondere werden im Rahmen der Subjekt-Kategorie "Datenbank" alle Entity-Typen mit ihren Attributen und Wertebereichen sowie alle Beziehungen zwischen den Entity-Typen, ebenfalls mit Attributen und Wertebereichen, beschrieben und dokumentiert. Prinzipiell können alle während des Design-Prozesses entwickelten Schemata, insbesondere also auch ein ER-Entwurf oder ein Funktionsmodell, das die Informationsverarbeitung eines Anwendungsbereichs aus funktionaler Sicht beschreibt, im Data-Dictionary abgelegt werden. Das gleiche gilt für die Transformationsregeln, die beim Übergang von einem Schema zum anderen Schema (beispielsweise bei der Übertragung des ER-Entwurfs in ein logisches Schema auf Basis des relationalen Modells) angewendet werden. Ferner können wichtige Projektdaten (Richtlinien, Konventionen, Dokumente, Netzpläne) vorgehalten werden. Das Data-Dictionary soll ebenfalls die Integrität der gespeicherten Informationen sicherstellen. Insgesamt soll im Idealfall der gesamte Lebenszyklus eines Informationssystems von der konzeptionellen Entwurfsphase bis zur Wartungsphase unterstützt beziehungsweise abgedeckt werden. Schließlich können im Data-Dictionary Privilegien, Zugriffsberechtigungen und -beschränkungen der Benutzer abgelegt werden. Gerade

dadurch wird Zugriffsschutz auf Feldebene möglich. Bei jedem Zugriff auf ein bestimmtes Attribut kann das Datenbanksystem über das Data-Dictionary feststellen, ob der Benutzer beziehungsweise das Anwendungsprogramm die entsprechende Berechtigung besitzt.

Die Subjekte der einzelnen Kategorien können nun wie gewöhnliche Entities zueinander in Beziehung gesetzt werden. Man erhält eine **Vernetzung** der Subjekte. Die grafische Darstellung (Abb. 45) erinnert stark an **Stücklistenstrukturen** und zeigt deutlich hierarchische Unter- beziehungsweise Überordnungsfunktionen auf.

In Abb. 45 ist ein einfaches Schema für ein Data-Dictionary abgebildet. Dabei bedeutet "c", daß ein übergeordnetes Subjekt ein untergeordnetes Subjekt **enthält** (contains). Entsprechend steht "m" für die **Manipulation** eines untergeordneten Subjektes durch ein übergeordnetes Subjekt. Die Bezeichnungen "p-k" beziehungsweise "f-k" drücken aus, daß eine Relation einen bestimmten Primär- beziehungsweise Fremdschlüssel besitzt

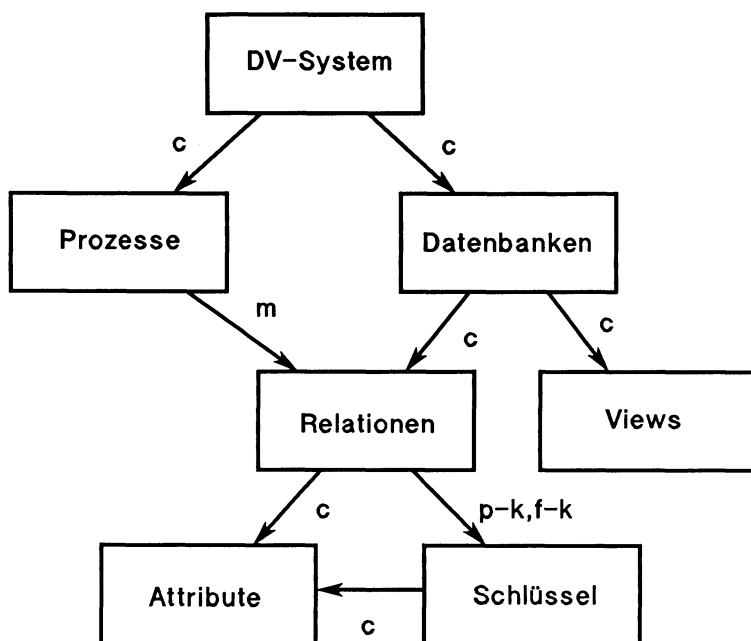


Abb. 45

Wie bei der bekannten Stücklistenverarbeitung kann man auch in einem Data-Dictionary alle untergeordneten Subjekte, die von einem übergeordneten Subjekt abhängen, bestimmen (**Auflösung, Explosion**). Sucht man nur die einem Subjekt direkt untergeordneten Subjekte, so spricht man von **einstufiger** Auflösung. Neben der einstufigen Auflösung gibt es auch die **mehrstufige** Auflösung, bei der untergeordnete Objekte ihrerseits weiter aufgelöst werden.