Vorstellung von zwei generischen Datenbankkonzepten zur Modellierung von technisch-wissenschaftlichen Datenbanken

D. Becker[1], U. Hussels[1], J. Nagel[1]

1 Einleitung

Die vielfältigen Fragestellungen und Forderungen im Ingenieurs- und Umweltbereich führten zu der Entwicklung zweier generischer Datenbankentwicklungswerkzeuge. Forderungen stellen dabei die schnelle und wirtschaftliche Anpassung eines Datenmodells - sowohl in der Entwicklungs- als auch der Produktionsphase - und damit einer Applikation an neue Anforderungen z.B. aufgrund neuer rechtlicher Regelungen der europäischen Kommision dar. Hiebei ist zu beachten, dass einmal die Institutionen durchgängig arbeitsfähig bleiben wie auch, dass Daten aus einem „Alt"-System ohne Probleme in ein neues Datenmodell übertragbar sein müssen. Basierend auf den bestehenden und immer stärker zunehmenden Bestimmungen auf EU-Ebene steht zudem die Zusammenführung unterschiedlicher Fachrichtungen, z.B. Luft- und Wasserbereich, immer mehr im Vordergrund, da diese gemeinsam in einer EU-Richtlinie berichtet werden müssen.

Die Beschreibung und ein folgender Vergleich der generischen Datenmodellierungsansätze soll eine Entscheidung über mögliche Einsatzbereiche ermöglichen.

2 Struktureller Aufbau der generischen Datenmodellierungsansätze

2.1 Struktureller Aufbau von UDIS-GEN

UDIS-GEN erfordert wegen seines objektorientierten Ansatzes für die Entwicklung eines fachlichen Datenmodells zunächst die Definition von Typen, denen Instanzen als individuelle Ausprägungen zugeordnet werden können. Eine Instanz wird über eine Instanz-ID als Individuum innerhalb eines Typs eindeutig identifiziert. Ein Typ ist eine allgemeingültige Beschreibung für eine Menge von ähnlichen (realen oder abstrakten) Gegenständen. Die Abgrenzung mehrerer Typen gegeneinander ist willkürlich und orientiert sich an der fachlichen Aufgabenstellung und deren Erfordernissen.

Ein Typ kann einen Elter-Typ besitzen. In diesem Fall besitzt jede Instanz des Typs auch genau eine Elter-Instanz des Elter-Typs. Über diesen Weg können Typ-Bäume definiert werden. Jeder Typ-Baum besitzt genau einen Typ ohne Elter-Typ (d. h. eine Wurzel).

Das erste Element des Konzeptes stellt somit die Typen-Hierarchie dar. Dazu parallel erfolgt die Einrichtung einer Instanzverwaltung. Zu jeder Instanz wird die Elter-Instanz (sofern vorhanden) mit abgelegt.

Im Allgemeinen besteht zwischen den Instanzen von Elter-Typ und Kind-Typ eine 1 : n-Beziehung. Es ist aber auch möglich, diese auf eine 1 : 1-Beziehung einzuschränken. Dadurch entstehen sogenannte „Paralleltypen". Obwohl zwischen den Instanzen der Paralleltypen eine 1 : 1-Beziehung besteht, erhalten die Instanzen beider Typen eigene Instanz-IDs. Die Elter-Instanzen der Instanzen des Paralleltyps sind die zugehörigen Instanzen des Basis-Typs.

Auf dieser Ebene der Datenverwaltung werden keine weiteren Informationen (d. h. Daten) zu den Typen und deren Instanzen verwaltet. Es ist auch nicht erforderlich, dass zu den Instanzen eines Typs überhaupt mehr als eine Instanz-ID verwaltet wird. In der Praxis

[1] RISA Sicherheitsanalysen GmbH, Krumme Str. 55, 10627 Berlin

stellen die Instanzen eines Typs ohne weitere Datenverwaltung schlicht einen Thesaurus dar.

Zu einer Instanz können ggf. historisch verwaltete Ausprägungen von Eigenschaften abgelegt werden. Mit der Definition eines Typs wird festgelegt, für welche Eigenschaften seiner Instanzen Daten abgelegt werden können. Die Eigenschaften selbst bestehen aus einer Eigenschafts-ID und einem Datentyp, dem die Ausprägung der Eigenschaft angehören muss. Eine Eigenschaft kann einem Typ nur einmal zugeordnet werden und je Instanz und Zeitpunkt mit maximal einer Ausprägung versehen werden, um in ein klassisches Tabellenwerk abgelegt werden zu können.

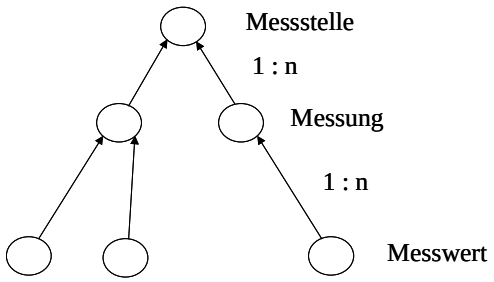

Abb. 2: Modell für UDIS-GEN

Relationen können ausschließlich über Elter-Kind Beziehungen von Typen definiert werden. Es besteht dabei die Möglichkeit, entweder 1 : n oder 1 : 1-Relationen anzulegen. Relationen in der Form n : m sind hier nur indirekt über Zeigereigenschaften von Subtypen modellierbar. Die Relationen der Form 1 : 1 (auch zwischen unterschiedlichen Typ-Hierarchien) werden mit Hilfe von Zeigereigenschaften definiert. Dazu existiert ein besonderer Datentyp „Zeiger auf Typ“, dessen mögliche Ausprägungen die Instanzen eines festgelegten Typs sind.

Im Vergleich zur Modellierung in RISA-GEN (s. Abb. 2) wird das Modell, wie in Abbildung 2 dargestellt, aufgebaut.

Es zeigt sich, dass in UDIS-GEN klare Typ-Hierachien abgelegt werden.

2.2 Struktureller Aufbau von RISA-GEN

Das Konzept für RISA-GEN ist ebenfalls objektorientiert und beruht auf einem universellen Prototypenmodell. Ein Prototyp wird durch eine Anzahl von Attributen gebildet, die die Eigenschaften der Instanzen dieses Prototypen beschreiben. Jede Instanz besitzt innerhalb des Prototypen eine eindeutige ID (Instanz-ID).

Ein Prototyp kann in Subprototypen eingeteilt werden. Subprototypen stellen eine „part of“-Beziehung zu Prototypen dar. Das generische Modell bietet dabei die Möglichkeit, beliebig viele solcher Subprototypeinteilungen vorzunehmen. Jeder Satz einer (nicht notwendigerweise vollständigen) Subprototypeinteilung eines Prototyps wird als Hierarchie bezeichnet und mit einer Hierarchie-ID abgelegt. Andere Prototypen können unter der selben Hierarchie-ID ebenfalls Subprototypeinteilungen bekommen, die zunächst nichts miteinander zu tun haben müssen.

Aus dieser Subprototypeinteilung wird eine Hierarchie, indem jedem Prototyp einer Instanz des Prototypen eine andere Instanz zugeordnet wird. Auf diesem Weg können Daten von einer tieferen Ebene (Subinstanzebene) einer höheren Ebene zugeordnet werden.

Beispiel 1:

Die Adresse ist ein Teil der Personaldaten. Andererseits können auch Institutionen Adressen haben. Folglich wird sowohl für den Prototyp der Personen als auch für den Prototyp der Institutionen ein Subprototyp „Adresse“ angelegt. Nun kann bei jeder Person und jeder Institution der Subprototyp „Adresse“ eine Instanz des neu einzuführenden Prototyps „Adressen“ zugewiesen werden. Die in dem Prototyp „Personen“ und „Institutionen“ angezeigten Adressdaten werden aus dem Prototyp „Adressen“ bezogen und nur dort physikalisch abgelegt.

Hierarchien können dabei n-stufig aufgebaut werden.

Zur Erstellung eines Datenmodells zur Erfassung von Messwerten, die an einer physikalisch bestehenden Messstelle ermittelt werden, kann bei RISA-GEN wie folgt vorgegangen werden (s. Abb. 2):

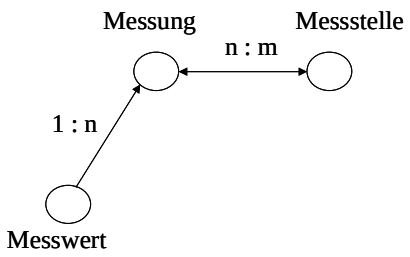

Abb. 2: Modell für RISA-GEN

Eine Messstelle erhält einen Zeiger auf eine Messung, die wiederum auf Messstellen zeigen kann. Der eigentliche Messwert wiedrum zeigt auf die Messung. Auf diese Art und Weise ist es möglich, Hierachien alternativ auch über Relationen aufzubauen.

3 Vergleich der generischen Ansätze

UDIS-GEN und RISA-GEN bieten beide jeweils ein Konzept zur Erstellung fachlicher Datenmodelle vor dem Hintergrund der Bereitstellung einer generischen Datenbank-Anwendungssoftware. Mit beiden Ansätzen lassen sich dieselben Inhalte, lediglich anders strukturiert abbilden. Welcher der Ansätze für eine Fragestellung geeigneter ist, hängt vermutlich weniger vom fachlichen Modell als von der Art und Weise des Zugriffs auf die Daten und der Intensität der Modelländerungen ab.

3.1 Gemeinsamkeiten der Ansätze

3.1.1 Allgemein

Allgemeine Gemeinsamkeiten können wie folgt benannt werden:

- Beide Ansätze basieren auf der Objektorientiertheit bei der Definition eines fachlichen Datenmodells.
- Bei beiden Ansätzen werden Typen (Prototypen) definiert, deren individuelle Ausprägungen als Instanzen bezeichnet werden.
- Bei beiden Ansätzen werden Eigenschaften definiert und den Typen zugeordnet, deren Ausprägungen die Instanzen individuell beschreiben.

3.2 Unterschiede der Ansätze

3.2.1 Definition von Ausprägungen

Bei der Definition von Ausprägungen in dem jeweiligen Ansatz kann ein entscheidender Unterschied festgestellt werden:

- In UDIS-GEN kann eine Eigenschaft für eine Instanz zu einem Zeitpunkt höchstens eine Ausprägung besitzen.
- RISA-GEN hingegen erlaubt grundsätzlich, dass gleichzeitig mehrere Ausprägungen existieren (bzw. eine geordnete Liste von Ausprägungen existiert).

3.2.2 Relationen-Definition für 1 : n-Verhältnisse

In UDIS-GEN können:

- 1 : n Relationen ausschließlich in einer Elter-Kind Beziehung zwischen Typen abgelegt werden. Damit befinden sich diese Relationen immer innerhalb der Typhierarchie eines sog. (Proto-) Typs.
- Die 1 : n-Relationen werden dadurch nicht im eigentlichen Datenteil, sondern in der Instanzverwaltung abgelegt, was einen Performance-Vorteil darstellen kann. Eine typische Typenhierarchie stellt die Kette <Messstelle – 1 : n – Messung – 1 : n – Messwert> dar.

In RISA-GEN können:

- beliebige 1 : n Relationen über Zeigereigenschaften mit Mehrfachausprägung realisiert werden. Diese können zwischen allen (Proto-) Typen (auch mit sich selbst) hergestellt werden. Diese Relationen werden im Datenteil abgelegt. Der Zugriff kann damit nicht schneller erfolgen, als auf andere dynamisch konfigurierte Eigenschaften.

Messstellen, Messungen und Messwerte würden in RISA-GEN als unabhängige Prototypen mit Zeigereigenschaften, die von jedem Messwert auf eine Messung und von jeder Messung auf eine Messstelle zeigen, konfiguriert. Damit entsteht auch kein Problem, wenn zu einem späteren Zeitpunkt plötzlich Messungen aufgenommen werden müssen, die nicht einer Messstelle, sondern mehreren Messstellen gleichzeitig zugeordnet sind (z. B. Mittelwerte über mehrere Messstellen von denen die Einzelwerte nicht verfügbar sind). In UDIS-GEN würde dies eine erhebliche Veränderung des Datenmodells erfordern, in RISA-GEN nur eine Umstellung einer Einfachausprägung auf Mehrfachausprägung für eine Zeigereigenschaft.

3.2.3 Relationen-Definition für 1 : 1-Verhältnisse

Die 1 : 1-Relationen können in UDIS-GEN als Sonderfälle der 1 : n-Relationen (Parallelty-pen) oder über Zeigereigenschaften zwischen beliebigen Typen realisiert werden.

In RISA-GEN bestehen diese beiden Möglichkeiten ebenfalls.

RISA-GEN bietet darüber hinaus die Möglichkeit zur Verwaltung von Subinstanzen. Subinstanzen sind normale Instanzen von beliebigen Prototypen. Innerhalb von RISA-GEN kann jedoch eine 1 : 1 Abbildungsvorschrift der Ausprägungen der Eigenschaften der Subinstanzen auf die Eigenschaften der Hauptinstanz abgelegt werden. Die Zahl der Subinstanzen einer Instanz wird mit der Definition des Prototyps, zu dem die Instanz gehört, festgelegt.

Diese Form der Instanz-Hierarchie, die sich von der Hierarchie in UDIS-GEN grundsätzlich unterscheidet, wird in der Instanzverwaltung abgelegt und ist daher performanter, als die „normale“ Relationenverwaltung in RISA-GEN.

Die Zuordnung von Subinstanzen kann rekursiv über beliebig viele Ebenen erfolgen. Das Konstrukt wird dazu verwendet, gleiche Teildefinitionen unterschiedlicher Prototypen mit einem gemeinsamen „Unter-“ Prototyp zu beschreiben und die Daten von diesem „Unter-“ Prototyp 1 : 1 auf die Instanzen der verschiedenen Prototypen der höheren Ebene

abzubilden. Dies kann die Datenverwaltung in bestimmten Fällen sehr vereinfachen und sorgt für völlige Redundanzfreiheit bei der physikalischen Datenablage.