

Aus dem Institut für Agrar- und Ernährungswissenschaften
(Geschäftsführender Direktor: Prof. Dr. K. Eder)

der Naturwissenschaftlichen Fakultät III
(Dekan: Prof. Dr. P. Wycisk)

der Martin-Luther-Universität Halle-Wittenberg



Hybride Modellierung operativer und analytischer Daten, dargestellt am Beispiel des Precision Dairy Farming

Dissertation

zur Erlangung des akademischen Grades
doctor agriculturarum (Dr. agr.)

vorgelegt von
Diplom-Agraringenieur Christian Schulze
geb. am 11.09.1976 in Luckenwalde

Gutachter: Prof. Dr. J. Spilke
Prof. Dr. P. Wagner
Prof. Dr. W. Lehner

Verteidigung am: 07.04.2008

Halle/Saale 2008

urn:nbn:de:gbv:3-000013737

[<http://nbn-resolving.de/urn/resolver.pl?urn=nbn%3Ade%3Agbv%3A3-000013737>]

Inhaltsverzeichnis

Verzeichnis der Abbildungen	IV
Verzeichnis der Abbildungen im Anhang	VIII
Verzeichnis der Tabellen	IX
Verzeichnis der Tabellen im Anhang	IX
Verzeichnis der Abkürzungen	X
1 Einleitung.....	1
2 Datenmanagement und Datenmodelle für betriebliche Informationssysteme.....	4
2.1 Datenmanagement in der Landwirtschaft.....	4
2.1.1 Bestandsaufnahme der Datenorganisation in der Landwirtschaft.....	5
2.1.1.1 Prozess- und Unternehmenszweigebene	5
2.1.1.2 Unternehmensebene	7
2.1.1.3 Datenaustausch und Anwendungen auf Ebene externer Organisationen	8
2.1.2 Anforderungen an eine betriebliche Datenorganisation.....	11
2.2 Konzepte des Datenmanagements in der Informatik.....	12
2.2.1 Operative Anwendungssysteme und deren Datenmodelle.....	12
2.2.2 Strategische Anwendungssysteme und deren Datenmodelle.....	14
2.2.2.1 On-Line Analytical Processing	15
2.2.2.2 Data Warehouse	17
2.3 Lösungsansatz der hybriden Datenmodellierung.....	19
3 Grundlagen der Datenmodellierung	22
3.1 Charakterisierung verschiedener Datensichten.....	22
3.1.1 Operative Datensicht	22
3.1.2 Analytische Datensicht.....	23
3.2 Modellierungsprozess	24
4 Informationsbedarfsanalyse	27
4.1 Grundlagen	27
4.1.1 Aufgabe der Informationsbedarfsanalyse.....	27
4.1.2 Analyseansätze zur Ermittlung des der Informationsbedarfs.....	28
4.1.3 Methoden der Informationsbedarfsanalyse	29
4.2 Informationsbedarfsanalyse im Rahmen des Precision Dairy Farming.....	30
4.2.1 Vorgehensweise und Durchführung.....	30
4.2.2 Ergebnisse der Informationsbedarfsanalyse.....	32
4.2.3 Diskussion der Informationsbedarfsanalyse.....	33

5 Semantischer Entwurf	35
5.1 Überblick über existierende Modellierungsansätze.....	35
5.1.1 Modellierungstechniken für operative Daten.....	35
5.1.2 Modellierungstechniken für analytische Daten.....	37
5.1.3 Aspekte der Zeitmodellierung im semantischen Entwurf.....	39
5.1.3.1 Aspekte der Zeitmodellierung im operativen Entwurf.....	39
5.1.3.2 Aspekte der Zeitmodellierung im analytischen Entwurf.....	41
5.2 Semantischer Datenbankentwurf im Rahmen des Precision Dairy Farming.....	42
5.2.1 Semantische Modellierung operativer Daten.....	42
5.2.1.1 Anwendungsspezifische Modellierungsanforderungen.....	42
5.2.1.2 Modellierungsnotation.....	42
5.2.1.3 Modellierungsgrundsätze.....	43
5.2.1.4 Ausgewählte Ergebnisse.....	47
5.2.2 Semantische Modellierung analytischer Daten.....	55
5.2.2.1 Anwendungsspezifische Modellierungsanforderungen.....	55
5.2.2.2 Modellierungsnotation.....	56
5.2.2.3 Ausgewählte Ergebnisse.....	58
5.2.2.4 Aggregationstypen von Kennzahlen.....	61
5.2.3 Diskussion der semantischen Datenbankentwürfe.....	64
6 Logischer Entwurf	66
6.1 Grundlagen des relationalen Modells.....	67
6.2 Abbildung operativer Daten im relationalen Modell.....	68
6.2.1 Standardisierte Transformationsregeln für E/RM-Elemente.....	68
6.2.2 Ausgewählte Ergebnisse der relationalen Modellierung operativer Daten.....	70
6.3 Abbildung analytischer Daten im relationalen Modell.....	78
6.3.1 Standardisierte Transformationsregeln für mE/RM-Elemente.....	78
6.3.2 Transformationsregeln für die mE/RM-Erweiterungen.....	79
6.3.3 Ausgewählte Ergebnisse der relationalen Modellierung analytischer Daten.....	81
6.4 Schemaintegration operativer und analytischer Daten.....	85
6.4.1 Identische Schemastrukturen.....	85
6.4.2 Verschiedene Schemastrukturen.....	86
6.4.3 Schemazusammenführung am Beispiel von Krankheitsdaten.....	88
6.5 Metamodell.....	90
6.5.1 Anforderungen an das Metamodell.....	90
6.5.2 Ausgestaltung des Metamodells.....	92
6.6 Diskussion des logischen Datenbankentwurfes.....	95

7 Aspekte des physischen Datenbankentwurfes.....	97
7.1 Aspekte der Konsistenzsicherung.....	97
7.2 Aspekte der Datenmodifikation.....	102
7.3 Undefinierte Intervallgrenzen zeitbezogener Gültigkeiten.....	104
7.4 Konsolidierte Sicht über mehrere Relationen.....	106
7.5 Aspekte der Indizierung.....	109
8 Schlussfolgerungen	112
9 Zusammenfassungen	118
9.1 Zusammenfassung	118
9.2 Summary.....	120
Literaturverzeichnis	122
Glossar	131
Anhang A: Referenzmodell Milcherzeugung.....	135
Anhang B: Material der Informationsbedarfsanalyse.....	144

Verzeichnis der Abbildungen

Abbildung 1: Schichtenmodell der Informationsverarbeitung in der Landwirtschaft (nach SPILKE 2003)	5
Abbildung 2: Nutzungsarten des Datenaustausches von Landwirtschaftsunternehmen	9
Abbildung 3: Datenmanagement operativer und analytischer Anwendungssysteme	12
Abbildung 4: Beispiel eines Datenwürfels für das Krankheitsauftreten bei Tieren.....	16
Abbildung 5: OLAP-Operationen Roll-up und Drill-down	17
Abbildung 6: OLAP-Operationen Pivotierung, Slice und Dice	17
Abbildung 7: Datenflüsse in Data Warehouse-Systemen	18
Abbildung 8: Klassifikationsschema und Klassifikationshierarchie.....	24
Abbildung 9: Schritte des Datenbankentwurfsprozesses	24
Abbildung 10: Informationsteilmengen (nach PICOT et al., 2001, S. 82).....	27
Abbildung 11: Prozessliste mit Merkmalsgruppen, Ausschnitt Milchgewinnung.....	32
Abbildung 12: Beispiel einer Attributversionisierung durch Entitätstypspaltung im E/RM.....	40
Abbildung 13: Versionisierung eines Beziehungstyps mittels Zeit-Entitätstyp im E/RM.....	41
Abbildung 14: Veränderungen in Klassifikationsschemata und Klassifikationshierarchien [Intervallanfang; Intervallende] – geschlossenes Intervall; [Intervallanfang; UC[- nach oben offenes Intervall (UC – until change)	42
Abbildung 15: E/RM-Basiselemente: Entitätstyp, Beziehungstyp, (Primärschlüssel-) Attribut und Kardinalitäten in min-max-Notation	43
Abbildung 16: Zusätzliche E/RM-Konstrukte: Schwacher Entitätstyp mit identifizierendem Beziehungstyp (links) und totale, disjunkte Generalisierung/Spezialisierung (rechts)	43
Abbildung 17: Verwendung von Typ-Entitätstypen zur Zusammenführung identischer Attribute.....	45
Abbildung 18: Ereignis als Beziehungstyp	45
Abbildung 19: Ereignis als schwacher Entitätstyp.....	45
Abbildung 20: Ausbildung von Schlüsselkaskaden	46
Abbildung 21: Nutzung einer ID zur Verhinderung von Schlüsselkaskaden	46
Abbildung 22: E/RM zum Tier	47

Abbildung 23: Hierarchien zwischen Organisationseinheiten	48
Abbildung 24: E/RM der Gültigkeiten von Organisationseinheiten	48
Abbildung 25: E/RM der Zuordnung des Tieres zu den Organisationseinheiten	49
Abbildung 26: E/RM der Zuordnung von Tiergruppen, Betriebsstätten, Ställen und Stallabteilen untereinander	49
Abbildung 27: E/RM der Zuordnung von Betriebsstätten und Abrechnungseinheiten zu Unternehmen	49
Abbildung 28: E/RM für Spezialisierung von Maßnahmen.....	51
Abbildung 29: E/RM für Maßnahmen	51
Abbildung 30: E/RM zur Abbildung von Krankheitsdiagnosen	52
Abbildung 31: E/RM zur Vererbung technischer Objekte.....	53
Abbildung 32: E/RM für Reinigung/Desinfektion.....	53
Abbildung 33: E/RM zu Fütterungsdaten des Einzeltieres	55
Abbildung 34: mE/RM-Elemente: Faktbeziehung mit Kennzahl, Klassifikationsstufe, Klassifikationsbeziehung (nach BAUER und GÜNZEL, 2004, S. 167).....	57
Abbildung 35: Erweiterungen der mE/RM-Notation: Eigenschaftsfenster und Kennzeichnung temporaler Restriktion (SCHULZE et al., 2005)	57
Abbildung 36: mE/RM der Dimension „Organisationseinheiten“ mit parallelen Klassifikationspfaden.....	58
Abbildung 37: mE/RM der Dimension „Zeit“ mit parallelen Klassifikationspfaden	59
Abbildung 38: mE/RM für Krankheitsbehandlungen	61
Abbildung 39: Notationselemente zur Darstellung relationaler Schemata	67
Abbildung 40: Relationale Transformation von Entitätstypen, Attributen und n:m Beziehungstypen	68
Abbildung 41: Relationale Transformation von 1:n-Beziehungstypen und schwachen Entitätstypen	69
Abbildung 42: Varianten der relationalen Transformation einer Spezialisierung/Generalisierung	70
Abbildung 43: Relationales Schema zum Tier.....	71
Abbildung 44: Relationales Schema der zeitlichen Gültigkeiten von Organisationseinheiten.....	72
Abbildung 45: Relationales Schema für Zugehörigkeiten des Tieres.....	73
Abbildung 46: Relationales Schema für Zugehörigkeiten von Tiergruppen, Stallabteilen und Ställen zu Betriebsstätten	73
Abbildung 47: Relationales Schema für Zugehörigkeiten von Betriebsstätten und Abrechnungseinheiten zu Unternehmen	73

Abbildung 48: Relationales Schema für die Spezialisierung einer Maßnahme	74
Abbildung 49: Relationales Schema für Maßnahmen.....	75
Abbildung 50: Relationales Schema für Diagnosen	75
Abbildung 51: Beispiel für die Ausprägung der Relationsschemata „diagnose“, „diagnosesystem“ und „zuordnung_extern“	76
Abbildung 52: Relationales Schema für die Spezialisierung von technischen Objekten	76
Abbildung 53: Relationales Schema für Reinigungen/Desinfektionen.....	77
Abbildung 54: Relationales Schema für die Fütterung von Einzeltieren.....	77
Abbildung 55: Relationale Transformation von Klassifikationsstufen (KS) und Klassifikationsbeziehungen	78
Abbildung 56: Relationale Transformation einer Faktbeziehung	78
Abbildung 57: Relationale Transformation der temporalen Restriktion einer Klassifikationsbeziehung	79
Abbildung 58: Relationale Transformation eines Eigenschaftsfensters für Kennzahlen.....	80
Abbildung 59: Relationales analyseorientiertes Teilschema der Dimension „Organisationseinheiten“ (Klassifikationsstufen Euterviertel, Tier, Tiergruppe, Stallabteil, Stall, Betriebsstätte)	82
Abbildung 60: Relationales analyseorientiertes Teilschema der Dimension „Organisationseinheiten“ (Klassifikationsstufen Tier, Abrechnungseinheit, Betriebsstätte, Unternehmen).....	82
Abbildung 61: Relationales analyseorientiertes Schema der Dimension „Zeit“	83
Abbildung 62: Beispiel für die Ausprägung des Teilpfades Zeitpunkt→Tag→Wochentag→Woche der Dimension „Zeit“	83
Abbildung 63: Relationales analyseorientiertes Schema für Krankheitsbehandlungen.....	84
Abbildung 64: Gegenüberstellung der operativ motivierten und der analyseorientiert motivierten Relation „tiergruppe“	86
Abbildung 65: Gegenüberstellung operativer und analyseorientierter Modellstrukturen der Zeit-Dimension	86
Abbildung 66: Gegenüberstellung operativer und analyseorientierter Schemastrukturen für Diagnosen	87
Abbildung 67: Gegenüberstellung operativer und analyseorientierter Schemastrukturen für Krankheitsbehandlungen	88
Abbildung 68: Gemeinsames relationales Schema (Ausschnitt) für Krankheitsdaten	89
Abbildung 69: Datenschema und multidimensionale Meta-Daten	91
Abbildung 70: E/RM des Metamodells zur Abbildung analyseorientierter Semantik.....	92

Abbildung 71: Relationales Schema des Metamodells zur Abbildung analyseorientierter Semantik.....	92
Abbildung 72: Beispiel für die Ausprägung des Metamodells	94
Abbildung 73: Sachlich inkonsistente Tabelle durch Zeitstempelung trotz Wahrung der Entitätsintegrität	100
Abbildung 74: Sachlich inkonsistenter Tabellenzustand durch Zeitstempelung trotz Anwendung der referentiellen Integrität	100
Abbildung 75: Beispiel für das Aktualisieren eines Objektes in einer Relation	104
Abbildung 76: Abfrage der am 1.5.2007 gültigen Version der Frischmelker- Tiergruppe auf die Tabelle „tiergruppe“ (Abbildung 74)	106
Abbildung 77: Integrierte Sicht auf Maßnahmen am Tier über eine VIEW- Definition in der PostgreSQL-Notation auf die Tabellen „massnahme“, „massnahme_tier“ und „vorfall_untersuchung“ (Abbildung 48, Abbildung 49).....	107
Abbildung 78: Funktionsdefinition in der PostgreSQL-Notation zum Einfügen einer Maßnahme an einem Tier in die Tabellen „massnahme“, „massnahme_tier“ und „vorfall_untersuchung“ (Abbildung 48).....	108
Abbildung 79: Struktur des Aufrufes der Funktion „fkt_massnahme_tier_einfuegen“ in der PostgreSQL-Notation zum Einfügen einer Maßnahme an einem Tier	108
Abbildung 80: SQL-Abfrage zur Bestimmung der Tiergruppen-Zugehörigkeit von Einzeltieren in der PostgreSQL-Notation	110
Abbildung 81: SQL-Abfrage zu Fütterungsdaten von Einzeltieren und Tiergruppen in der PostgreSQL-Notation.....	110
Abbildung 82: Daten- und Konsistenzebenen eines Data Warehouse-Systems (nach LEHNER, 2003, S. 23, 163)	115

Verzeichnis der Abbildungen im Anhang

Abbildung A 1: Darstellungselemente des Referenzmodells.....	135
Abbildung A 2: Referenzmodell, Teil Organisationseinheiten.....	136
Abbildung A 3: Referenzmodell, Teil Stammdaten Tier und Stammdaten Abrechnungseinheit.....	137
Abbildung A 4: Referenzmodell, Teil Ereignisse Tier.....	138
Abbildung A 5: Referenzmodell, Teil Fütterung	139
Abbildung A 6: Referenzmodell, Teil Gesundheit.....	140
Abbildung A 7: Referenzmodell, Teil Milchgewinnung	141
Abbildung A 8: Referenzmodell, Teil Technik.....	142
Abbildung A 9: Referenzmodell, Krankheitsbehandlungen, analytisch motivierte Strukturen	143
Abbildung A 10: Referenzmodell, Teil Metamodell	143

Verzeichnis der Tabellen

Tabelle 1: Durchschnittliche Anzahl an Tagen mit Milchsperr.....	24
Tabelle 2: Matrix der Analyseansätze und -methoden für die Informationsbedarfsanalyse	30
Tabelle 3: Anzahl der Merkmale und Merkmalsgruppen je Teilverfahren und Organisationseinheit (SCHULZE et al., 2004)	33
Tabelle 4: Merkmalsliste für Fütterungsdaten als Beispiel für Strukturanalogien	54
Tabelle 5: Aggregierbarkeit von Kennzahlen entsprechend ihres Aggregationstyps (nach LENZ und SHOSHANI [1997] und LEHNER [2003, S. 72], erweitert)	63
Tabelle 6: Zuordnung von Aggregationstypen zur statistischen Merkmalskalierung.....	64
Tabelle 7: Semantik der Datenmodifikationsoperationen in Gültigkeitszeit- Datenbanken in Objekt- und Objektversions-Betrachtungsweise	103

Verzeichnis der Tabellen im Anhang

Tabelle A 1: Beschriftung im Referenzmodell	135
Tabelle A 2: Datentypen des Referenzmodell	135
Tabelle A 3: Teilverfahren, Prozesse und Merkmalskomplexe der Informationsbedarfsanalyse	144
Tabelle A 4: Merkmale und Merkmalsgruppen der Informationsbedarfsanalyse	146

Verzeichnis der Abkürzungen

ADAPT	Application Design for Analytical Processing
ADIS	Agricultural Data Interchange Syntax
ADED	Agricultural Data Element Dictionary
ADR	Arbeitsgemeinschaft Deutscher Rinderzüchter e.V.
agroXML	agarspezifisches XML-Datenformat
ANSI/SPARC	American National Standards Institute / Standards Planning And Requirements Committee
bzw.	beziehungsweise
BCS	Body-Condition-Score
CASE	Computer Aided Software Engineering
DBMS	Datenbankmanagementsystem
DFM	Dimensional Fact Model
DLG	Deutsche Landwirtschafts-Gesellschaft e.V.
EDV	Elektronische Datenverarbeitung
E/RM	Entity-Relationship-Modell
eE/RM	erweitertes Entity-Relationship-Modell
ETL	Extraction-Transformation-Load
FK	Foreign Key - Fremdschlüssel
GIS	Geografisches Informations-System
HIT	Herkunftssicherungs- und Informationssystem für Tiere
ID / id	Identifikationsnummer
InVeKoS	Integriertes Verwaltungs- und Kontrollsystem
IT	Informations-Technologie
KTBL	Kuratorium für Technik und Bauwesen in der Landwirtschaft e.V.
KS / ks	Klassifikationsstufe
LKV	Landeskontrollverband
LOM	Lebensohrmarken-Nummer
mE/RM	multidimensionales Entity-Relationship-Modell
MLP	Milchleistungsprüfung
MML	Multidimensional Modeling Language
mUML	multidimensional Unified Modeling Language
NULL	NULL-Wert eines Attributs
ODS	Operational Data Store

OLAP	On-Line Analytical Processing
OLTP	On-Line Transaction Processing
PK	Primary Key - Primärschlüssel
SERM	strukturiertes Entity-Relationship-Model
SQL	Eigenbegriff, jedoch üblicherweise mit "Structured Query Language" erklärt
TO / to	technisches Objekt
UML	Unified Modeling Language
vgl.	Vergleiche
z.B.	zum Beispiel

1 Einleitung

Die Landwirtschaft war und ist, wie wohl jeder Wirtschaftszweig, starken Veränderungen unterworfen. Als Nahrungsmittel- und Rohstofflieferant kam ihr in der Vergangenheit durch einen politisch forcierten „Bestandsschutz“ mittels Subventionierung und Binnenmarktschutz jedoch eine Sonderstellung zu. Nicht zuletzt aufgrund einer Überproduktion an Nahrungsmitteln in den Industrieländern und der finanzpolitischen Notwendigkeit von Ausgabenreduzierungen der öffentlichen Haushalte wird diese Sonderstellung von der Gesellschaft zunehmend in Frage gestellt. Gleichzeitig wird in verstärktem Maße die Sicherheit von Nahrungsmitteln thematisiert. Darin begründen sich für die Landwirtschaft zwei wesentliche wirtschaftliche Rahmenbedingungen:

- Ein zunehmender Wettbewerb aufgrund von Marktliberalisierung und Subventionskürzungen erfordert eine Effizienzverbesserung der Produktion.
- Die wachsende gesellschaftliche Forderung nach Transparenz der Produktion und Rückverfolgbarkeit von Produkten, gekoppelt an eine erhöhte Erwartungshaltung in Bezug auf Umwelt- und Tierschutz, erfordert angepasste Produktions- und Informationskonzepte zur Absatzsicherung landwirtschaftlicher Produkte.

Der Notwendigkeit einer Effizienzsteigerung kann dabei nur zum Teil über neue oder verbesserte Produktionsverfahren begegnet werden. Einen immer größeren Stellenwert nimmt dagegen die Information als Produktionsfaktor ein (DOLUSCHITZ und BAUR, 1997, S. 16 f.). Dies gilt umso mehr bei der Realisierung von Transparenz und Rückverfolgbarkeit in der Produktion.

Die Erfüllung beider Teilaufgaben, Produktionsprozessoptimierung und -transparenz nach außen, begründet die Notwendigkeit betrieblicher Informationssysteme. Ausgangspunkt hierfür bilden Daten aus Produktions- und Geschäftsprozessen, die in hinreichender Weise zur Verfügung stehen müssen. Eine entsprechend große Bedeutung kommt dem Datenmanagement zu.

Die rasante Entwicklung der Prozesstechnik in der Landwirtschaft der letzten Jahre und Jahrzehnte verbesserte die Möglichkeiten der Datenerfassung, lässt jedoch bisher die grundlegende, prozessübergreifende und langfristige Datenverfügbarkeit im Sinne eines

angepassten Datenmanagements weitgehend vermissen. Besonders die Milcherzeugung als wichtiger Zweig der Tierproduktion weist hier noch erhebliche Defizite auf. In der Folge ist beispielsweise die Milcherzeugung durch eine Reihe von Insellösungen zur Datenspeicherung gekennzeichnet. Eine zu fordernde prozessübergreifende Datennutzung zur Aufwertung erfasster Daten zu Informationen und damit zur Unterstützung von Entscheidungsprozessen findet nur in geringem Maße statt. Gerade darauf basieren jedoch die Konzepte des Precision Livestock Farming bzw. seine Spezialisierung auf das Milchrind Precision Dairy Farming. Dessen Definition als „...integrativer Ansatz für eine nachhaltige Erzeugung von Milch mit gesicherter Qualität sowie einem hohen Grad an Verbraucher- und Tierschutz“ (SPILKE et al., 2003, S. 19) unterstreicht den interdisziplinären und integrierenden Charakter. Grundlage der Umsetzung dieses Bewirtschaftungsansatzes als Systemkonzept kann jedoch nur eine bislang fehlende integrierte und konsistente Datenbasis auf Betriebsebene sein.

Zielsetzung

Abgeleitet aus den kurz skizzierten Defiziten im Datenmanagement und der Datenverfügbarkeit ist es das Ziel der Arbeit, einen Vorschlag zur Datenmodellierung als wesentlichen Bestandteil des Datenmanagements und damit die Basis betrieblicher Informationssysteme für den Bereich der Milcherzeugung zu erarbeiten. Besonderes Augenmerk ist hierbei auf die simultane Beachtung sowohl operativer Daten zur kurzfristigen als auch analytischer Daten zur strategischen Entscheidungsunterstützung zu legen. Basierend auf einer Analyse des Informationsbedarfs sind relevante Daten zu identifizieren und entsprechend der Nutzungsrichtung im Datenmodell in geeigneter Weise abzubilden. Der Fokus ist dabei nicht nur auf unternehmensinterne, produktionsprozessorientierte Daten zu legen, sondern auch auf den Informationsbedarf von Kunden, Lieferanten und anderen externen Organisationen im Rahmen von Geschäftsprozessen und Kooperationsverpflichtungen. Dabei soll speziell der Frage nachgegangen werden, inwieweit operative und analytische Datensichten in einem parallelen Modellierungsprozess mit einem gemeinsamen relationalen Modell als Ergebnis berücksichtigt werden können. Das erstellte Datenmodell soll die Funktion eines Referenzmodells¹ für den Bereich der Milcherzeugung erfüllen.

¹ „Ein Referenzmodell ist eine Spezifikation eines Metamodells für einen bestimmten Anwendungsbereich mit dem Anspruch einer gewissen Allgemeingültigkeit. [...] Ein Referenzmodell ist auf Wiederverwertung [...] ausgerichtet.“ (LASSMANN et al., 2001, S.372)

Aufbau der Arbeit

Die Arbeit gliedert sich im Wesentlichen in drei wesentliche Teile:

- **Bestandsaufnahme:** Eine Bestandsaufnahme der Datenorganisation in der Landwirtschaft und der Datenmodellierung für betriebliche Informationssysteme in der Informatik identifiziert bestehende Konzepte, Defizite und Potentiale. Daraus wird folgend ein Lösungsansatz abgeleitet (Kapitel 2).
- **Grundlagen:** Als Voraussetzung für die Umsetzung des Lösungsansatzes sind zum Verständnis wichtige Grundlagen zu erläutern (Kapitel 3). Das betrifft insbesondere die Unterscheidung operativer und analytischer Daten und den Modellierungsprozess selbst.
- **Datenmodellierung:** Die Darstellung und Diskussion der Datenmodellierung zur Umsetzung des Precision Dairy Farming – Konzeptes erfolgt in den Kapiteln 4 bis 6. Dabei werden jedem Kapitel theoretische Grundlagen zum jeweiligen Modellierungsschritt vorangestellt. Abschließend werden Aspekte des physischen Datenbankentwurfes behandelt (Kapitel 7).

2 Datenmanagement und Datenmodelle für betriebliche Informationssysteme

Ausgehend von gegenwärtigen und zukünftigen Anforderungen werden Ansprüche, Potentiale, Defizite und die derzeitige Organisation im Datenmanagement der Landwirtschaft beschrieben. Ein Blick in die Informatik soll eine Bestandsaufnahme außerlandwirtschaftlicher Lösungsansätze liefern und damit eine Einordnung der Aufgabenstellung in existierende Konzepte ermöglichen.

Die klassische Teilung in prozess- und auswertungsorientierte Informationssysteme mit entsprechenden logischen Datenmodellen ist bei Beachtung der vorliegenden Aufgabenstellung kritisch zu hinterfragen. Als Lösungsansatz für die notwendige grundlegende Neugestaltung der Datenorganisation in der Milcherzeugung wird eine hybride Datenmodellierung unter Berücksichtigung prozess- und auswertungsorientierter Aspekte vorgestellt.

2.1 Datenmanagement in der Landwirtschaft

Die betriebliche Datenorganisation in der Landwirtschaft, speziell in der Milcherzeugung ist durch eine Vielzahl von historisch gewachsenen Datenerfassungs- und Datenhaltungssystemen gekennzeichnet. Eine Integration bestehender Systeme erfolgt größtenteils nur innerhalb der Produktfamilie eines Herstellers, z.B. zwischen Prozessrechentechnik und Herdenmanagementprogramm mit teilweise erheblichen Unzulänglichkeiten in der Datenqualität (AMMON und SPILKE, 2006). Ansätze einer weitergehenden Prozessintegration beschränken sich weitestgehend auf Kommunikationsaspekte (AIDS/ADED [LKV NORDRHEIN-WESTFALEN, 2004; RATSCHOW, 2004, S. 24 ff.]) und haben ihren derzeitigen Fokus außerhalb der Milcherzeugung (agroXML [DOLUSCHITZ und KUNISCH, 2004; KTBL, 2005; MARTINI et al., 2006]). Unabhängig davon fehlt auf Betriebsebene jedoch weiterhin eine prozessübergreifende, anwendungsunabhängige Datenbasis wie sie AUERNHAMMER schon 1988 forderte (AUERNHAMMER, 1988). Diese sollte detaillierte Prozessdaten für andere Anwendungen zur Verfügung stellen und gleichzeitig die Funktion einer langfristigen Datenspeicherung für Analysezwecke übernehmen. Der Nutzen einer solchen konsolidierten Datenbasis wird zum einen in Synergieeffekten bei der Entscheidungsunterstützung

bisher autonom arbeitender Prozesse gesehen, die ohne prozessübergreifende Datenintegration nicht nutzbar wären. Zum anderen ermöglicht die langfristig konsistente Datenspeicherung Zeitreihenanalysen sowohl zur Optimierung einzelner Prozesse als auch ganzer Verfahrenabschnitte bzw. zur Bewertung des Gesamtverfahrens.

2.1.1 Bestandsaufnahme der Datenorganisation in der Landwirtschaft

Ausgangspunkt zur Eingrenzung von Untersuchungsgegenstand und Lösungsansatz soll eine Bestandsaufnahme der Datenorganisation innerhalb der Landwirtschaft, mit dem jedoch nicht ausschließlichen Fokus auf der Milchproduktion bilden. Zur Strukturierung werden hierbei die organisatorischen Ebenen der Informationsverarbeitung herangezogen (Abbildung 1; SCHÖN, 1993, S. 23 f.; SPILKE, 2003).

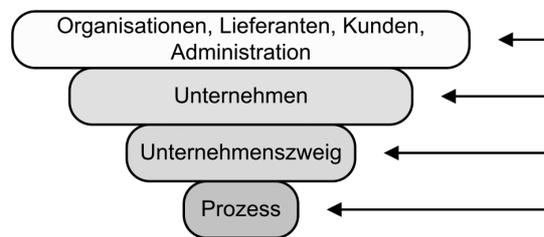


Abbildung 1: Schichtenmodell der Informationsverarbeitung in der Landwirtschaft (nach SPILKE 2003)

2.1.1.1 Prozess- und Unternehmenszweigebene

Vielfältige Forschungsergebnisse und technische Entwicklungen der letzten Jahre auf Ebene der Prozessrechentchnik bilden den wesentlichen Auslöser für weiterführende Fragestellungen zur Datenorganisation und -nutzung in der Landwirtschaft. Als entscheidend ist dabei die Entwicklung neuartiger Sensoren und deren Übertragung in praktische Anwendungen zu sehen.

Schlüsselsysteme im Bereich des Pflanzenbaus stellen Positionsbestimmungssysteme und Sensorik zur kontinuierlichen Pflanzenbestandscharakterisierung dar (RATSCHOW, 2004, S.8). Ziel ist die ortspezifische Erfassung von Bestandsmerkmalen zur Steuerung, Kontrolle und Bewertung eines ortspezifischen Produktionsmitteleinsatzes.

Im Bereich der Tierproduktion gilt die Einzeltiererkennung als Schlüsseltechnologie (SCHLÜNSEN, 1988; KLINDTWORTH, 1999, S. 6; RATSCHOW, 2004, S. 8). Im Zusammenhang damit wurden verschiedene Sensorsysteme (Bestimmung von Milchmengen und Milchqualität, Futteraufnahme, Lebendmasse, Aktivität, etc.) entwickelt, um einzeltierbezogene Daten zu bestimmen und so die Grundlage zur automatisierbaren einzeltier-

bezogenen Erfassung von physiologischen Parametern und Leistungen geschaffen. Hier liegt das Ziel in der Steuerung, Kontrolle und Bewertung des Produktionsmitteleinsatzes auf Ebene einer kleinstmöglichen Produktionseinheit, in diesem Fall auf Ebene des Einzel-tieres.

Automatisierte und computergestützte operative Entscheidungsprozesse sind in vielfältiger Weise Gegenstand der Forschung und Entwicklung (z.B. HARPAIN, 1990; DEVIR et al., 1993; SEIFERT, 1993; KLEINHENZ et al., 1998; DE MOL, 2000). Neben aktuellen Prozessdaten spielen dabei auch vergangenheitsbezogene Daten bei der Anwendung von Methoden zur Unterstützung operativer Entscheidungsprozesse eine Rolle (z.B. AMMON und SPILKE, 2005; ENGLER et al., 2005).

Die Datenorganisation auf Unternehmenszweigebene als Basis operativer Aufgaben ist jedoch in der wissenschaftlichen Literatur unterrepräsentiert, obwohl sie Datenverfügbarkeit und Datenqualität maßgeblich beeinflusst und zum Teil gravierende Defizite aufweist (AMMON und SPILKE, 2006).

Eine grundlegende Aufarbeitung des Themas im deutschsprachigen Raum entstand Ende der 1980er Jahre unter Federführung der Deutschen Landwirtschafts-Gesellschaft e.V. (DLG) sowohl für den Bereich der Tier- als auch der Pflanzenproduktion (DLG, 1985; DLG, 1987). Augenmerk wurde hier jedoch auf die funktionalen Anforderungen der Datenverarbeitung in der Landwirtschaft gelegt. Weiterführende Arbeiten beschäftigten sich mit dem Integrationsgedanken auf Betriebsebene (DLG, 1990), beschränkten sich jedoch aufgrund von Realisierungsmöglichkeiten auf eine Anwendungs- statt auf eine Datenintegration (KUHLMANN et al., 1990). Spätere Arbeiten führten zu einem Datenmodell für Marktfruchtbetriebe (DLG, 1992), das in diesem Sinne ein Teilmodell für den Unternehmenszweig der Pflanzenproduktion enthält. Für den Bereich der Tierproduktion sind keine entsprechenden Ergebnisse bekannt.

Aktuelle Ansätze einer den Anforderungen angepassten Datenorganisation liegen vor allem aus dem Bereich des Precision Farming vor. Aufgrund der starken Prägung der Organisation ortsbezogener Daten durch Geografische Informations-Systeme (GIS) mit proprietären Dateiformaten spielen Datenbanken zum großen Teil nur zur Metadatenverwaltung (BILL und KORDUAN, 2004, S. 29) und zur Datenspeicherung einzelner Anwendungsmodulare eine Rolle (BÖTTINGER et al., 2004, S. 6, 10; SCHWAIBERGER, 2004). Die Notwendigkeit einer konsistenten, zentralen GIS-erweiterten Datenbasis ist dennoch

erkannt und als Beispielapplikation umgesetzt (LINSEISEN et al., 2000; LINSEISEN, 2002, S. 50, 81).

Für den Bereich der Milcherzeugung sind neuere Arbeiten zur Datenorganisation aus den Niederlanden (BRAND et al., 1995) und weitgehend unveröffentlichte Projekte der Universität Kiel (KARSTEN et al., 2005; KARSTEN und STAMER, 2006) sowie der Landesanstalt für Landwirtschaft Sachsen bekannt. Die letztgenannten Arbeiten haben ihren Fokus auf der Forschungsunterstützung durch Datenintegration angeschlossener Versuchsbetriebe. Entsprechend liegen ihre Aufgaben hauptsächlich in einem direkten Anwendungsbezug. Die Datenschemata der implementierten Forschungsdatenbanken wurden bisher nicht offen gelegt. Die methodisch ausgerichtete Arbeit von BRAND et al. (1995) trennt ein Informationsmodell in Prozess- und Datenmodell. Neben einer Beschreibung der Prozess-Teilmodelle erfolgt jedoch lediglich die Veröffentlichung der Grobstruktur des Datenmodells anhand der Teilprozessmodelle. Eine Offenlegung der Datenstruktur im Sinne eines Referenzmodells ist damit nicht gegeben.

2.1.1.2 Unternehmensebene

Gegenstand computergestützter Systeme auf Unternehmensebene ist die Unterstützung

- wertorientierter Abrechnungsaufgaben,
- Berichts-, Kontroll- und Analyseaufgaben sowie
- strategischer Planungs- und Entscheidungsaufgaben.

Die Grundlage zur Realisierung dieser Aufgaben bilden vor allem Daten der mengenorientierten operativen Systeme und externer Datenquellen zur Charakterisierung der Unternehmensumwelt, die in hinreichender Weise zur Verfügung stehen müssen. Die unterschiedlichen Verwendungsziele erfordern eine datenseitige Kopplung der Anwendungen (DOLUSCHITZ, 2002), um Datenredundanz zwischen verschiedenen Anwendungssystemen auf ein Minimum zu beschränken. Entsprechend besteht die Forderung nach integrierten Anwendungssystemen.

Ansätze mit dem Anspruch integrierter Informations- oder Managementsysteme auf Unternehmensebene existieren sowohl als Spezialisierung für Milchvieh- (PIETERSMA et al., 1998) als auch für Marktfruchtbau-Unternehmen (BLÖNNIGEN, 1990; DLG, 1992). SCHWECKE UND WALTER (1990) sowie PEDERSON et al. (1997) veröffentlichten Lösungen für Gemischtbetriebe. Bis auf PIETERSMA et al. (1998) beschränken sich die genannten Arbeiten jedoch weitestgehend auf wertorientierte Abrechnungsaufgaben ohne Beachtung strategischer Analyse-, Planungs- und Entscheidungsaufgaben.

Dem Datenmodell als Bestandteil von Informationssystemen widmen nur BLÖNNIGEN (1990), DLG (1992) und die bereits angeführte Arbeit von BRAND et al. (1995) Beachtung, während sich die anderen publizierten Arbeiten auf eine Funktionsintegration konzentrieren. Die Rolle von Datenbanken zur Datenintegration wird dagegen bereits von GRAF (1990) beschrieben.

Losgelöst vom Integrationsgedanken existieren seit langem verschiedene strategische Expertensysteme zur Entscheidungsunterstützung auf Unternehmensebene (z.B. BERG und SCHMIDT-PAULSEN, 1990; WAGNER und KUHLMANN, 1991).

Des Weiteren wird zur Unterstützung von Analyseaufgaben im Rahmen eines unternehmensweiten Managements eine flexible Auswertungsunterstützung propagiert, die jedoch im Bereich der Landwirtschaft bisher nur konzeptionellen Charakter erreicht (SCHIEFER, 2002, S. 338 ff.; WENDT und DOLUSCHITZ, 2002, S. 189 f.). WENDT und GEORG (2001) verweisen explizit auf das Data Warehouse-Konzept als Lösungsansatz.

2.1.1.3 Datenaustausch und Anwendungen auf Ebene externer Organisationen

Landwirtschaftsunternehmen sind durch einen hohen Grad an Geschäftsbeziehungen untereinander und zu anderen Unternehmen und Organisationen wie Zucht- und Kontrollverbänden, Kunden, Lieferanten, öffentliche Verwaltungen etc. gekennzeichnet. Entsprechende Beziehungen sind dabei immer an einen unterschiedlich intensiven Austausch von Daten gebunden. Besondere Bedeutung erhält der Datentransfer durch aktuell forcierte Entwicklungen im Bereich von Qualitätssicherungssystemen, transparenter Produktion, Rückverfolgbarkeit von Nahrungsmitteln und damit landwirtschaftlicher Rohstoffe. Wesentlich verbindendes Merkmal dieser Entwicklungen ist der Kerngedanke der Datenakkumulation und Datenbereitstellung zur Befriedigung eines Informationsbedürfnisses der Handelspartner, des Endverbrauchers oder des Landwirtes selbst.

Prinzipiell kann in produktionsunterstützende, dokumentierende und administrative Nutzungsarten des Datenaustausches von Landwirtschaftsbetrieben unterschieden werden, deren Grenzen jedoch fließend sind (Abbildung 2).

Die bisher bedeutendste Aufgabe eines Datenaustausches ist in der Produktionsunterstützung zu sehen. Der Schwerpunkt liegt hierbei vor allem auf einem Datenfluss parallel zu physischen Transaktionen als Produkt- und Dienstleistungsbeschreibung (produktbegleitende Daten) und in einer Datenbereitstellung als Produkt (Daten als Ware).

An Relevanz gewinnen jedoch Ansätze des Outsourcing von informationstechnischen Aufgaben (WENDT et al., 2004), wobei insbesondere die Datenverarbeitung als Dienst-

leistung (LUDOWICY et al., 2002, S. 118) Fragen des Datenaustausches betrifft. Dazu zählen ausdrücklich auch Unternehmensvergleiche und Beratungsangebote auf Ebene von regionalen oder nationalen Zucht-, Kontroll- oder Beratungsverbänden (WIGGANS, 1994; SWENSSON und SEDERBLAD, 1997; REENTS et al., 2003; SCHULZE et al., 2007). Ebenfalls in diese Kategorie ist der an eine Datenverarbeitung für Analysezwecke gekoppelte Datenaustausch einzuordnen (z.B. für rechenintensive Entscheidungsunterstützungsmodelle).

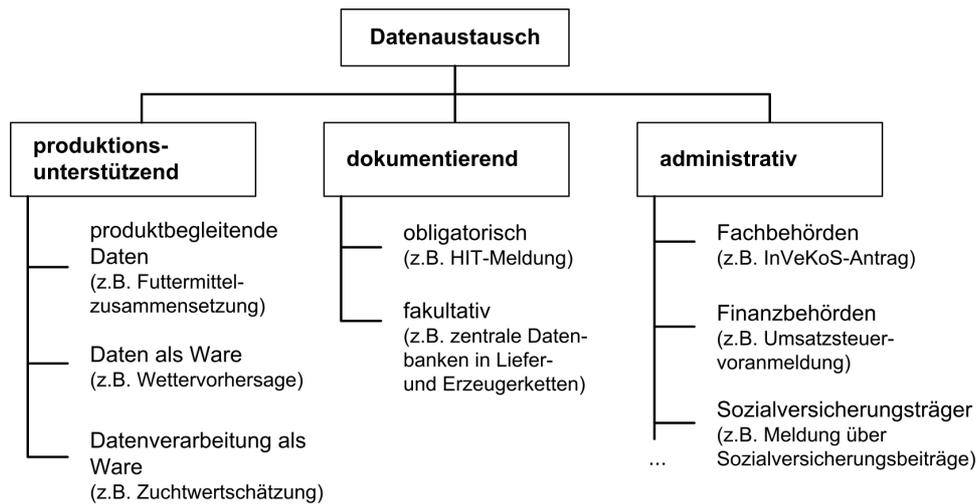


Abbildung 2: Nutzungsarten des Datenaustausches von Landwirtschaftsunternehmen

Eine zweite Säule des Datenaustausches ist in der Unterstützung primär dokumentarischer Aufgaben zu sehen. Eine Differenzierung nach der Motivation des Datenaustausches führt zu einer Unterscheidung in einen gesetzlich verordneten und einen prinzipiell freiwilligen Datenaustausch. Dabei kommt dem Datenaustausch im Rahmen von Liefer- und Erzeugerketten (Supply Chains) als prinzipiell freiwilligem Datenaustausch² eine wachsende Bedeutung zu. Hintergrund ist die Etablierung zentraler Datenbanken zur Befriedigung produktspezifischer Informationsbedürfnisse entlang von Liefer- und Erzeugerketten (BEERS, 2002; HANNUS et al., 2003). Es ist jedoch zu klären, inwieweit eine zentrale Dokumentation von Supply Chains Aufgaben der unternehmenseigenen Datenorganisation übernehmen muss (vgl. BODMER und HORVÁTH, 2002). Die Abgrenzung zum produktbegleitenden Datenaustausch ist motivationsbedingt und äußert sich in einer zentralen Datenspeicherung zur Sicherung der Rückverfolgbarkeit. Als Beispiele zum gesetzlich verordneten Datenaustausch zu Dokumentationszwecken sind die Meldungen laut Viehverkehrsverordnung bei der HIT-Datenbank (Herkunftssicherungs- und Informationssystem für

² Der Datenaustausch im Rahmen von Liefer- und Erzeugerketten wird als freiwillig bezeichnet, weil die Mitgliedschaft in entsprechenden Ketten nicht zwingend ist, der Datenaustausch als Folge der Mitgliedschaft unter Umständen jedoch schon.

Tiere) oder die Meldungen zur Erfüllung der Offenlegungspflichten gegenüber dem Handels- und Genossenschaftsregister anzuführen.

Eine dritte, davon zu unterscheidende Art des Datenaustausches ist in der Unterstützung administrativer Aufgaben zu sehen. Zu nennen sind hier vor allem der Datenverkehr mit Fachbehörden (z.B. zur Agrarantragstellung), mit Finanzbehörden (z.B. zur Umsatzsteuervoranmeldung) und mit Sozialversicherungsträgern (z.B. Meldung sozialversicherungspflichtiger Beschäftigungsverhältnisse). Im Unterschied zum Datenaustausch mit rein dokumentierender Funktion betreffen die administrativ motivierten Austauschbeziehungen Daten, die die Grundlage von Verwaltungsaufgaben mit daran gekoppelten Zahlungsansprüchen und -vorgängen bilden.

Voraussetzung für jede Art des Datenaustausches ist zum einen die Standardisierung von Austauschprotokollen bzw. -sprachen (TAURUS [ATC, 1993]; ADIS/ADED [LKV NORDRHEIN-WESTFALEN E.V., 2004]; ISOagriNET [BAUFÖRDERUNG LANDWIRTSCHAFT GMBH, 2005]; agroXML [DOLUSCHITZ und KUNISCH, 2004; KTBL, 2005]). Zum anderen hängt der Erfolg maßgeblich von der unternehmensseitigen Bereitstellung geforderter Daten bzw. der Möglichkeit der Einordnung externer Daten in das unternehmensseitige Datenmanagement ab. Berührt sind davon alle Ebenen der Informationsverarbeitung eines Unternehmens, von der Prozess- bis zur Unternehmensebene. Entsprechend sind sowohl detaillierte, operative als auch aggregierte Daten Gegenstand eines Datenaustausches.

Unabhängig vom Datenaustausch existieren auf Ebene externer Organisationen Informationssysteme, die das betriebliche Anwendungsfeld tangieren. So stellten SWENSSON und SEDERBLAD (1997) eine Data Warehouse-Lösung für die Beratung von Milchviehbetrieben auf Basis der Milchleistungsprüfung vor. Mit der Herde als kleinster abgebildeter Einheit und der Fokussierung auf die überbetriebliche Beratung als Anwendungsfeld sind Ergebnisse jedoch nicht auf die betriebliche Einzeltierebene übertragbar.

Zur Erstellung von Informationssystemen für Nutztierpopulationen wurde mit APIIS (Adaptable platform independent information system) ein Framework veröffentlicht, das grundlegende Strukturen zur Speicherung einzeltierbezogener Daten offen legt. Das Projektziel besteht hauptsächlich in der Unterstützung des Entwurfes von Informationssystemen für Zuchtorganisationen und überregionale Genombanken. Veröffentlichte Arbeiten sind daher ausschließlich dieser Domäne zuzuordnen (FISCHER et al., 2002; GROENEVELD,

2004, GROENEVELD et al., 2004). Eine Adaption auf betriebliche Anforderungen ist nicht bekannt, aufgrund allgemeingültiger Datenstrukturen jedoch diskussionswürdig.

2.1.2 Anforderungen an eine betriebliche Datenorganisation

Die Bestandsaufnahme der Datenorganisation und der zu unterstützenden Aufgaben in der Landwirtschaft zeigt den Bedarf einer umfassenden Datenverfügbarkeit auf allen Ebenen der Informationsverarbeitung. Die Nutzung landwirtschaftlicher Prozessdaten zur Entscheidungsunterstützung und Dokumentation ist Stand der Wissenschaft bzw. Praxis. Die Verfügbarkeit qualitativ hochwertiger Ausgangsdaten wird dabei (meist stillschweigend) vorausgesetzt. Die Realisierung dieses Anspruches erfordert jedoch ein angepasstes Datenmanagement, dessen zentrales Element eine integrierte, historisierte, anwendungsunabhängige Datenbasis für alle Ebenen der Informationsverarbeitung der Landwirtschaft sein muss. Die folgenden Punkte sollen diese Forderung unterstreichen bzw. deren Notwendigkeit zusammenfassen:

- Die Planung, Steuerung und Kontrolle von Produktionsprozessen erfordern die Verfügbarkeit widerspruchsfreier und hinreichend detaillierter Prozessdaten. Eine dezentrale Datenhaltung auf Prozessebene mit der bekannten Gefahr von widersprüchlich gespeicherten Daten kann diese Anforderung nicht erfüllen.
- In Hinblick auf eine nötige Dokumentation von Produktionsprozessen sind Prozessdaten langfristig zu speichern. Unter Beachtung der Forderung einer zentralen Datenbasis kommt diese Aufgabe der Unternehmenszeigebene als unterste prozessübergreifende Ebene zu.
- In der vielgestaltigen Datennutzung, sowohl auf operativer als auch auf strategischer Aufgabenebene, begründet sich die Forderung nach einer anwendungsunabhängigen Datenbasis. Nur damit ist eine langfristig konsistente Datenverfügbarkeit unabhängig von verwendeten Anwendungssystemen gesichert.
- Die Unterstützung operativer und strategischer Aufgaben erfordert die Verfügbarkeit vergangenheitsbezogener Daten.
- Das Fehlen einer offen gelegten Datenstruktur für den Bereich der Milcherzeugung ist ein wesentlicher Nachteil bei der Entwicklung und Diskussion verschiedener Anwendungskonzepte und -systeme.

2.2 Konzepte des Datenmanagements in der Informatik

Die für die Milcherzeugung angeführten Anforderungen und Defizite wurden und werden prinzipiell auch in anderen Wirtschaftsbereichen diskutiert. Hier hat sich eine verwendungszweckbedingte Trennung in operative Administrations- und Dispositionssysteme (transaktionsorientiert) und strategische Planungs- und Kontrollsysteme (auswertungsorientiert) herausgebildet (LASSMANN et al., 2001, S. 480 ff.; STAHLKNECHT und HASENKAMP, 2005, S. 326 ff.). Diese Aufteilung ist auch in den verwendeten Datenmanagementkonzepten und den dahinter liegenden Datenmodellen wieder zu finden (Abbildung 3).

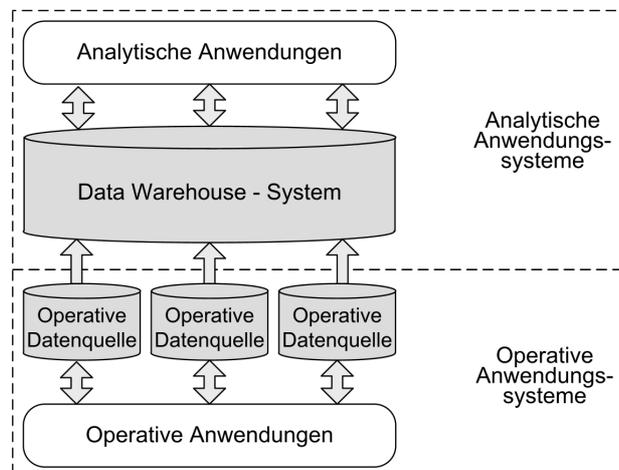


Abbildung 3: Datenmanagement operativer und analytischer Anwendungssysteme

Das nachfolgende Kapitel stellt notwendige Begriffe operativer und analytischer Informationssysteme und deren Merkmale vor. Dabei wird speziell auf das Konzept des Data Warehouse als integrierende, auswertungsorientierte Datenbasis und das Anwendungskonzept On-Line Analytical Processing (OLAP) eingegangen.

2.2.1 Operative Anwendungssysteme und deren Datenmodelle

Operative Anwendungssysteme stellen in heutiger Zeit betriebswirtschaftliche Standardsoftware für eine Vielzahl von Anwendungsbereichen dar (CHAMONI und GLUCHOWSKI, 1999a, S. 10) und sollen deshalb im Folgenden nur kurz behandelt werden.

Gegenstand operativer Aufgaben ist die Unterstützung kurzfristiger Entscheidungen zur direkten Umsetzung der Unternehmensziele (SCHWICKERT und FISCHER, 1996, S. 13 f.). Operative Entscheidungen beruhen auf der Kenntnis des aktuellen Zustands der entscheidungsrelevanten Umwelt. Entsprechende Anwendungssysteme dienen der Verwaltung und Bearbeitung von aktuellen Massendaten bei einer Vielzahl zugriffsberechtigter Nutzer, man denke an Kassensysteme in Kaufhäusern oder die Auftragsverwaltung in Versand-

häusern. Nutzer operativer Systeme sind vor allem die untere und mittlere Führungsebene in Unternehmen.

Der datenseitige Aspekt operativer Anwendungssysteme lässt sich zusammenfassend durch folgende Punkte beschreiben:

- eine hohe Aktualität und damit eine hohe Änderungshäufigkeit der Daten
- große Datenmengen
- kleine, klar abgegrenzte Realitätsausschnitte
- bis auf Ausnahmen fehlender Bedarf an historischen Daten

Die sich daraus ergebenden Anforderungen bezüglich der zu verwendenden Formen der Datenorganisation prädestinieren eine Datenbanknutzung als Basis operativer Anwendungssysteme (ZEHNDER, 2005, S. 18; BISKUP, 1995, S. 2 ff.; LASSMANN et al., 2001, S. 232 ff.; KEMPER und EICKLER, 2004, S. 17 f.). Als Datenmodell verbreiteter Datenbankmanagementsysteme hat sich das relationale Modell (CODD, 1970) durchgesetzt.

Die Zugriffscharakteristik operativer Anwendungssysteme stellt sich wie folgt dar (LEHNER, 2003, S. 18; BAUER und GÜNZEL, 2004, S. 9):

- einfache Anfragestruktur
- kleine Ergebnismengen von Anfragen
- transaktionale Zugriffe: Lesen, Schreiben, Aktualisieren, Löschen
- kurze Zugriffszeiten
- hohe Nutzerzahl
- hohe Zugriffshäufigkeit

Der charakteristische schreibende Zugriff auf Datenbanken zwingt zu einer Vermeidung von Redundanz, das heißt einer Mehrfachspeicherung desselben Datums. Andernfalls bestünde die Gefahr inkonsistenter Datenbankzustände aufgrund von Einfüge-, Aktualisierungs- und Löschanomalien. Deshalb zeichnen sich Datenbanken operativer Anwendungssysteme durch einen hohen Normalisierungsgrad aus. Daraus resultierende komplexe, detaillierte Datenmodelle mit transaktionaler Optimierung und weitestgehend fehlendem Zeitbezug sind Kennzeichen von Datenbanken dieser Anwendungsebene. In Abgrenzung zum nachfolgend behandelten analyseorientierten On-Line Analytical Processing kennzeichnen die angeführten Merkmale das On-Line Transactional Processing (OLTP) auf operativer Ebene.

Die verbreitete Beschränkung operativer Datenbanken auf die Abbildung aktueller Zustände ohne Berücksichtigung historischer Konstellationen des betrachteten Realitätsausschnittes ist für eine Reihe von Anwendungen unbefriedigend. Diesem Anwendungsbereich widmen sich Konzepte temporaler Datenbanken als Bestandteil operativer Informationssysteme (TANSEL et al., 1993; ELMASRI und NAVATHE, 2002, S. 799 ff.; KNOLMAYER und MYRACH, 1996; MYRACH, 2005). Hierbei liegt der Schwerpunkt auf der Abbildung von Gültigkeits- und Transaktionszeiten durch vorhandene und erweiterte Datenmodelle und Datenmanipulationssprachen.

2.2.2 Strategische Anwendungssysteme und deren Datenmodelle

Eine weitere Klasse von Anwendungssystemen stellen Führungssysteme dar. Ihnen sind die Funktionen der strategischen Planung und Kontrolle zugeordnet. Damit sind sie als Werkzeuge zur Bestimmung und Überwachung von Unternehmenszielen eindeutig von den operativen Anwendungssystemen mit der Aufgabe zur Unterstützung der Umsetzung von Unternehmenszielen abzugrenzen. Nutzergruppe ist demgemäß auch die mittlere und obere Führungsebene.

Sowohl für die strategische Planung als auch für die Kontrolle der Zielerreichung werden unternehmensinterne Informationen benötigt. Aufgabe strategischer Anwendungssysteme ist deshalb die adäquate Bereitstellung führungsrelevanter Daten auf Grundlage unternehmenseigener und externer operativer Daten (LASSMANN et al., 2001, S. 488; STAHLKNECHT und HASENKAMP, 2005, S. 331).

Die Unterschätzung strategischer Aufgaben erfordert darüber hinaus die Betrachtung von Unternehmensdaten unterschiedlicher Verdichtungsstufen in Zeitreihen und unterschiedlichen Zusammenhängen, entsprechend unterschiedlicher Hypothesen innerhalb von Unternehmensanalysen. Gekennzeichnet sind derartige Aufgaben durch weitestgehend unstrukturierte Lösungsansätze im Gegensatz zu standardisierten Aufgaben im operativen Umfeld. Gefordert ist deshalb eine gewisse Interaktivität durch flexible Softwaresysteme und eine adäquat flexible Datenspeicherung mit eindeutiger Analyseorientierung. Damit können analytische Informationssysteme wie folgt charakterisiert werden:

- Integration verschiedener operativer Datenquellen
- analyseorientierte Datenspeicherung
- komplexe Anfragestruktur
- flexible Anfragegegenstände
- große Ergebnismengen von Anfragen
- weitestgehende Beschränkung auf Lesezugriffe

- Tolerierung höherer Anfragezeiten
- kleine Nutzergruppe
- wenige Zugriffe pro Zeiteinheit

Zur Erfüllung derartiger Anforderungen wurden und werden in Informatik und Wirtschaftsinformatik Konzepte unter den Begriffen Data Warehouse als analyseorientierte Datenbasis und On-Line Analytical Processing als ein Vertreter von Analysetechnologien diskutiert, auf die im Folgenden näher eingegangen werden soll.

2.2.2.1 On-Line Analytical Processing

Unter dem Begriff des On-Line Analytical Processing werden Konzepte zur interaktiven Analyse multidimensionaler Datenbestände zusammengefasst. Als Definition gelten 12 von CODD et al. (1993) aufgestellte Anforderungen an OLAP-Systeme (CHAMONI und GLUCHOWSKI, 1999b, S. 232 ff.; LASSMANN et al., 2001, S. 496 ff.; BAUER und GÜNZEL, 2004, S. 98 ff.). Diese wurden unter dem Akronym FASMI (Fast Analysis of Shared Multi-dimensional Information) auf 5 Bewertungsregeln reduziert (PENDSE, 2005, BAUER und GÜNZEL, 2004, S. 101 f.) und sollen hier zur Abgrenzung des OLAP dienen:

- Geschwindigkeit: Nutzeranfragen sind schnell zu beantworten.
- Analyse: Dem Nutzer sind intuitiv bedienbare Analysefunktionen bereitzustellen, über das OLAP-Werkzeug selbst oder über externe, darauf zugreifende Produkte.
- Mehrbenutzerbetrieb: OLAP-Systeme müssen einen Mehrbenutzerbetrieb ermöglichen und daran angepasste Sicherheits- und Rechtekonzepte aufweisen.
- Multidimensionalität: Dem Nutzer muss eine multidimensionale konzeptuelle Datensicht geboten werden, unabhängig von der verwendeten Datenbanktechnologie.
- Information: OLAP-Systeme müssen den Zugriff auf alle von der Anwendung geforderten Daten ermöglichen.

Wie die Bezeichnung FASMI bereits nahe legt, definieren die OLAP-Kriterien folglich einen schnellen und mehrbenutzerfähigen analyseorientierten Zugriff auf umfassende mehrdimensionale Datenbestände.

Kerngedanke des OLAP ist damit die Multidimensionalität von Daten, die mit einer Strukturierung in qualifizierende und in quantifizierende Daten einhergeht, entsprechend Auswertestrukturen und Auswertungsgegenstand. Der Nutzer erhält dadurch eine, den jeweiligen Fragestellungen angepasste Sicht auf Daten (CODD et al., 1993; CHAMONI und GLUCHOWSKI, 1999a).

Zur Veranschaulichung wird gern eine Würfelmetapher herangezogen. Die Würfelzellen repräsentieren den Auswertungsgegenstand, z.B. betriebswirtschaftliche Kennzahlen, während die Würfelkanten den Analysekontext, den Bezugsrahmen der Kennzahlen, über so genannte Dimensionen darstellen (Abbildung 4).

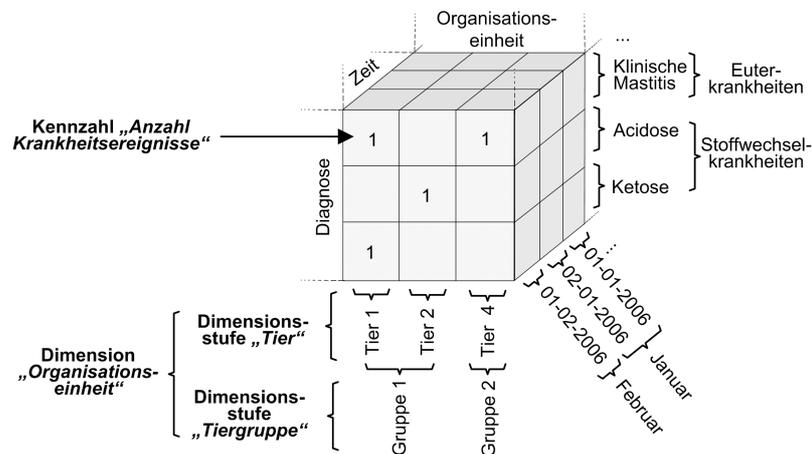


Abbildung 4: Beispiel eines Datenwürfels für das Krankheitsauftreten bei Tieren

Im Rahmen von Analysen kann es sinnvoll sein, die Dimensionen einer Kennzahl mithilfe existierender Hierarchien weiter zu gliedern, zu klassifizieren. Entsprechend können Klassifikationsstufen³ definiert sein, anhand derer eine Kennzahl disaggregiert bzw. aggregiert werden kann. Das führt zu Klassifikationspfaden⁴, die Auswertungsstrukturen für mögliche Analysen darstellen. Aggregation und Disaggregation von Kennzahlen entlang von Klassifikationspfaden als typische OLAP-Operationen werden auch als Roll-Up und Drill-Down bezeichnet (Abbildung 5). Weitere charakteristische Operationen sind Pivotierung (das Vertauschen von Dimensionen als Perspektivwechsel), Drill-Across (ein Wechsel des Datenwürfels oder eines Klassifikationspfades) sowie Slice (Reduktion einer Dimension auf eine Klassifikationsstufe – das Herausschneiden einer Scheibe aus dem Datenwürfel) und Dice (Reduktion auf einen Teilwürfel) (Abbildung 6). Ziel der Bereitstellung derartiger Operationen ist eine möglichst intuitive Navigation für den Nutzer im Datenbestand auf einer vordefinierten Struktur, um flexibel auf neue Fragestellungen reagieren zu können. Daraus ergeben sich jedoch Anforderungen an die Datenspeicherung, die im Folgenden dargestellt werden.

³ synonym: Klassifikationsebenen; Dimensionsstufen, -ebenen; Aggregationsstufen, -ebenen

⁴ synonym: Aggregationspfade; Auswertepfade

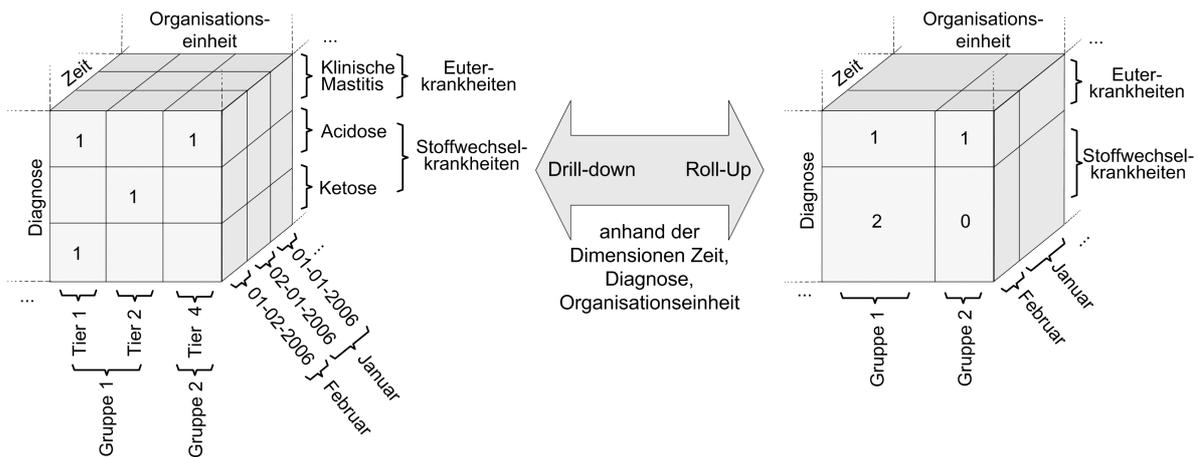


Abbildung 5: OLAP-Operationen Roll-up und Drill-down

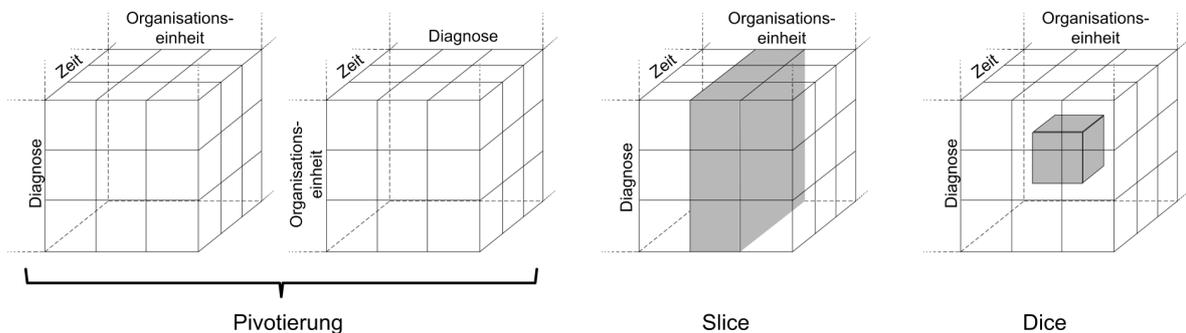


Abbildung 6: OLAP-Operationen Pivotierung, Slice und Dice

2.2.2.2 Data Warehouse

Analyseorientierte Anwendungen wie die im vorhergehenden Kapitel beschriebenen OLAP-Systeme erfordern den schnellen Zugriff auf große, unternehmensweite Datenmengen mit unterschiedlichem Aggregationsniveau. Operative, hoch normalisierte, auf transaktionale Datenmanipulationen ausgerichtete Datenbanken können diese Anforderungen nur unzureichend erfüllen. Als ein Lösungsansatz werden so genannte Data Warehouses gesehen, die explizit als Datenhaltungskonzept für analytische Informationssysteme dienen sollen. Definiert ist ein Data Warehouse dabei als eine themenbezogene, integrierte, zeitvariierende, nicht flüchtige Sammlung von Daten zur Unterstützung von Management-Entscheidungsprozessen (INMON, 1996, S. 33).

Das Data Warehouse stellt in analyseorientierten Datenorganisationskonzepten das zentrale Datenbanksystem dar (Abbildung 7). Datenquellen des Data Warehouse sind operative Datenbanken oder sonstige unternehmensinterne und externe Datensammlungen.

Als zentrale Analyse-Datenbank werden jedoch hohe Anforderungen an die Datenqualität gestellt. Es wird sogar der Anspruch eines „Single point of truth“ erhoben (DINTER et al.,

1998; WINTER und STRAUCH, 2004). Aus diesem Grund kommt dem Befüllen der Data Warehouse-Datenbank, dem Prozess des Extrahierens, Transformierens und Ladens (ETL) von Quelldaten eine besondere Bedeutung zur Sicherung der Datenqualität zu. Dabei werden relevante Daten aus den Quellsystemen extrahiert, in die Datenstrukturen des Data Warehouse transformiert, auf Einhaltung von Integritätsbedingungen überprüft und letztendlich in das Data Warehouse geladen. Die zentrale Datenspeicherung eines Data Warehouse ist jedoch nicht zwangsläufig auf eine Datenbank beschränkt, sondern kann aus mehreren Teil-Datenbanken bestehen, die, wie in Abbildung 7 dargestellt, beispielsweise Daten verschiedener Verarbeitungsschritte im Data Warehouse-Prozess speichern.

Neben einer zentralen, weitestgehend unternehmensweiten Data Warehouse-Architektur existieren Ansätze zur Etablierung so genannter Data Marts, die als thematisch abgegrenzte Teilsichten von Data Warehouses mit der Aufgabe einer spezialisierten analytischen Datenbasis für kleinere Unternehmenseinheiten (z.B. Fachabteilungen) dienen sollen. Der Vorteil liegt zum einen in einer schnelleren Datenverfügbarkeit bei lokal vorhandenen, replizierten Daten und zum anderen in der Möglichkeit der Optimierung auf spezifische Analyseaufgaben.

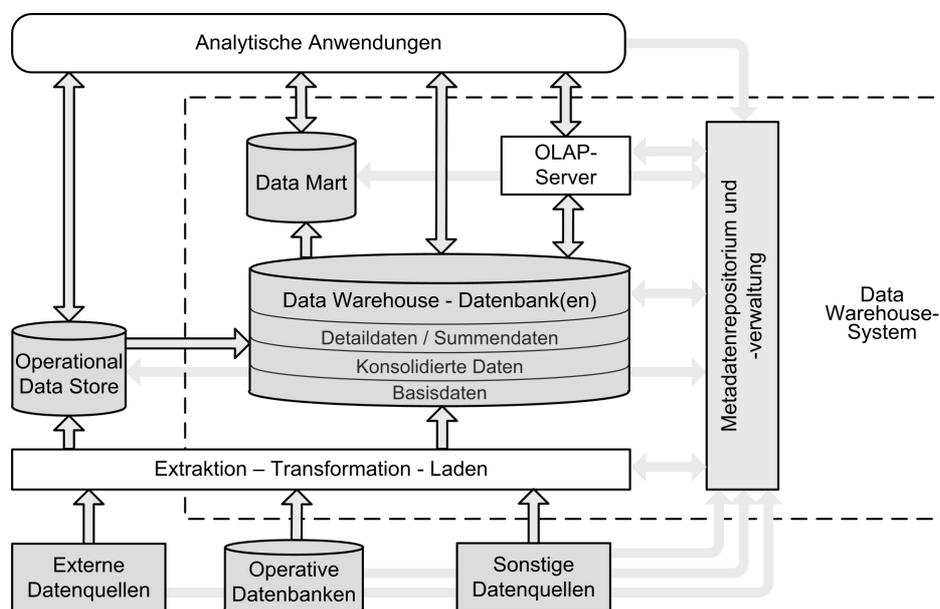


Abbildung 7: Datenflüsse in Data Warehouse-Systemen

Analytische Anwendungen greifen entweder direkt oder über Middleware (z.B. OLAP-Server) auf Data Warehouses und vorhandene Data Marts zu. Von entscheidender Bedeutung bei derartig integrativen Konzepten ist die Bereitstellung von Metadaten über alle Ebenen der Datenaufbereitung und Anwendung. Nur das Vorliegen vollkommener Infor-

mationen zu angebotenen Daten und deren Quellen kann die Grundlage von Analysefunktionalitäten bilden.

Erweitert wird das Konzept des Data Warehouse durch die Einführung von Operational Data Stores (ODS), die dem Bedürfnis der Managementebene nach aktuellen Geschäftsdaten in multidimensionaler Form entspringen. Dabei handelt es sich um einen, dem eigentlichen Data Warehouse vorgeschalteten und thematisch begrenzten Datenspeicher, der in kürzeren Abständen als das Data Warehouse selbst mit multidimensional aufbereiteten transaktionalen Daten operativer Systeme aktualisiert wird. Die durch einen ETL-Prozess schon bereinigten Daten des Operational Data Store können in einem nachgelagerten Schritt direkt oder mit einem verkürztem ETL-Prozess in das Data Warehouse gelesen werden. Das ursprüngliche Konzept sieht dabei keine Historisierung der ODS-Daten vor (KIMBALL, 1997; INMON, 1999, S. 13; KEMPER und LEE, 2003), während KIMBALL (1997) eine im Umfang beschränkte Historisierung vorschlägt („gleitender Monat“). Da der Operational Data Store ähnliche Datenstrukturen wie das eigentliche Data Warehouse aufweist, kann der Zugriff in adäquater Weise erfolgen.

Die vorangehend beschriebene Data Warehouse-Architektur stellt Data Warehouse-Umgebungen als Systeme mit mehreren Datenebenen heraus. Diese Datenebenen stellen jeweils verschiedene Sichten auf einen Realitätsausschnitt mit unterschiedlichen Integrations- und Aggregationsniveaus dar. Dabei notwendige Aktualisierungsprozesse, insbesondere ihre zeitliche Abfolge innerhalb und zwischen den Datenebenen bilden den Auslöser für Fragestellungen zur Konsistenz der Daten verschiedener Datenebenen eines Data Warehouse-Systems. Bei einer horizontalen Datenintegration innerhalb der Ebenen und der vertikalen Datenaktualisierung und -aggregation zwischen den Ebenen ist deshalb auf eine Konsistenzsicherung zu achten (LEHNER, 2003, S. 162 ff.). Diesem Aspekt kommt im Weiteren eine wichtige Rolle zu.

2.3 Lösungsansatz der hybriden Datenmodellierung

Ausgehend von den Anforderungen, die durch die Milcherzeugung an die Datenorganisation gestellt werden (vgl. Kapitel 2.1.2), erscheint eine prozessübergreifende Datenmodellierung im Sinne einer Neuorganisation der Datenhaltung unumgänglich. Das trifft insbesondere auf die Modellierung operativer Daten zu. Parallel besteht jedoch die

Forderung einer Unterstützung strategischer Aufgaben, die in der Informatik durch Konzepte des Data Warehousing als Analyseplattform angeboten werden. Dabei hat sich die Trennung in operative und analytische Informationssysteme mit eigenen Datenmodellen heraus gebildet. Als wesentliche Gründe sind zu nennen:

- [a] Die Etablierung operativer und analytischer Anwendungssysteme und entsprechender Datenmodelle verläuft größtenteils zeitlich getrennt. Das heißt, analytische Anwendungssysteme werden nachträglich zu operativen Systemen, jedoch auf ihnen aufbauend in die Unternehmens-IT eingeführt.
- [b] Die geforderte Verfügbarkeit vergangenheitsbezogener Daten in analytischen Informationssystemen führt zu Unterschieden in der Datenstruktur und der Handhabbarkeit (Abfragen, Datenmanipulationen) zwischen analyseorientierten, historisierenden und operativen Datenschemata.
- [c] Das unterschiedliche Nutzerverhalten beider Anwendungswelten zwingt zu unterschiedlichen Optimierungsmaßnahmen in Bezug auf das Zugriffsverhalten der Datenbankimplementierungen. Bei gemeinsamer Nutzung eines Datenbanksystems würden sich darüber hinaus beide Nutzungsarten gegenseitig im Laufzeitverhalten behindern.

Trotz dieser Unterschiede bilden beide Datensichten dennoch prinzipiell den gleichen Kontext ab, jedoch mit unterschiedlicher Granularität und verändertem Fokus, wodurch die Etablierung entsprechend getrennter Datenschemata ein gewisses Maß an Redundanz nach sich zieht.

Für die Neukonzeption der Datenorganisation im Rahmen des Precision Dairy Farming relativieren sich die Gründe für eine Trennung beider Modellwelten. Die unter [a] genannte zeitliche Trennung entfällt. Der vorliegende Anwendungsfall eröffnet die Möglichkeit beide Datensichten zeitgleich zu modellieren. Die Forderung der Verfügbarkeit vergangenheitsbezogener Daten zur Dokumentation und Unterstützung operativer Entscheidungen im Precision Dairy Farming (vgl. Kapitel 2.1.2) zwingt zur Historisierung des operativen Datenschemas. Damit entfällt auch der unter [b] genannte Unterscheidungsgrund operativer und analytischer Datenschemata, da sowohl vom analytischen als auch vom operativen Datenschema die Abbildung vergangenheitsbezogener Daten verlangt werden.

Die unter [c] angeführten Optimierungsunterschiede beider Datenbankimplementierungen spielen auf Ebene des semantischen und des logischen Datenbankentwurfes keine Rolle, so

dass eine Trennung beider Datenschemata auf konzeptueller Ebene nicht begründet werden kann. Dementsprechend wird als Lösungsansatz eine hybride Datenmodellierung als gemeinsamer Datenbankentwurf operativer und analytischer Daten mit dem Ziel eines gemeinsamen logischen Datenschemas angesehen.

Gestützt wird der Ansatz durch das Konzept des Operational Data Store. Dessen Rolle als Integrationsplattform operativer Daten zur Analyseunterstützung aktueller oder begrenzt historischer transaktionaler Daten rückt den hybriden Modellierungsansatz konzeptionell in die Nähe von Operational Data Stores und zeigt die notwendige Verknüpfung der operativen und der analytischen Ebene. Auch durch das OLAP-Konzept ist lediglich eine Trennung beider Datensichten auf konzeptueller Ebene begründet (CODD et al., 1993), nicht jedoch zwangsläufig auf logischer oder physischer Ebene. Es ist jedoch bisher kein Ansatz zur gemeinsamen Modellierung beider konzeptueller Sichten auf logischer Ebene bekannt. Lediglich die Ableitung multidimensionaler Strukturen, ausgehend von operativen Entity-Relationship-Modellen, ist beschrieben (GOLFARELLI et al., 1998a; BOEHNLEIN und ULBRICH-VOM ENDE, 1999) und kann damit als Vorstufe eines hybriden Modellierungsansatzes betrachtet werden.

Der hybride Modellierungsansatz wird im Folgenden als ein konzeptueller Modellierungsprozess operativer und analytischer Daten verstanden, bestehend aus semantischem und logischem Entwurf, der im Ergebnis ein gemeinsames logisches Datenschema liefert, das sowohl die Anforderungen einer operativen als auch analytischen Datensicht erfüllt. Dabei ist es unerheblich, wie die folgende physische Implementierung organisiert ist. Diese wird im Weiteren nicht als Bestandteil des hybriden Modellierungsansatzes gesehen. In diesem Sinne kann das logische Datenschema eines hybriden Modellierungsprozesses die Basis für die Implementierung

- einer integrierten operativen Datenbank,
 - einer konsolidierten Data Warehouse-Datenbasis und/oder
 - damit verbundener Operational Data Stores darstellen.
-

3 Grundlagen der Datenmodellierung

Das Verständnis für den in der vorliegenden Arbeit gewählten Lösungsansatz erfordert eine Einführung in die Methoden und Konzepte des Datenbankentwurfes. Grundlegend ist hierbei die ausdrückliche Unterscheidung in eine operative und eine analytische Datensicht als Basis eines hybriden Modellierungsansatzes. Das Kapitel arbeitet Unterschiede und Gemeinsamkeiten beider Datensichten heraus und verweist auf Konsequenzen für die Modellierung. Weiterhin wird der Prozess des Datenbankentwurfes kurz vorgestellt.

3.1 Charakterisierung verschiedener Datensichten

3.1.1 Operative Datensicht

Ausgehend von den Merkmalen operativer Anwendungssysteme (vgl. Kapitel 2.2.1) wird im Folgenden die Sicht auf Daten im operativen Umfeld genauer spezifiziert. Die Aktualität operativer Daten zur Charakterisierung der entscheidungsrelevanten Umwelt wurde bereits herausgearbeitet. Eine operative Datensicht muss deshalb notwendigerweise den aktuellen Zustand des abzubildenden Realitätsausschnittes wiedergeben, der über Zustände darin enthaltener Objekte und deren Beziehungen zueinander charakterisiert wird. Die Zustandsbeschreibung wiederum erfolgt über die Zuweisung von Eigenschaften. Der Begriff des Objektes bezeichnet in diesem Zusammenhang zum einen materielle Objekte (z.B. Gegenstände und Lebewesen) und zum anderen immaterielle Objekte (z.B. Unternehmen oder Konzepte). Ereignisse im Realitätsausschnitt sind bei dieser Sichtweise als Zustandsänderung der modellierten Objekte und Beziehungen zu sehen.

Ein Beispiel für ein operatives Datum ist die Frage nach einer bestehenden Milchsperrung für eine Milchkuh. Der Milchkuh kann unter anderem die Eigenschaft „Enddatum Milchsperrung“ zugewiesen werden. Für eine Kuh wird somit beispielsweise nach jeder Krankheitsbehandlung definiert, ob und bis wann die betreffende Kuh mit einer Milchsperrung belegt ist. Vor jedem Melken ist somit zu prüfen, ob aktuell eine Milchsperrung vorliegt und die Kuh deshalb nicht in die Ablieferungsmilch gemolken werden darf.

Die operative Datensicht stellt demnach originäre, zeitaktuelle und detaillierte Daten bereit. Damit im Zusammenhang steht eine hohe Änderungshäufigkeit in Abhängigkeit der Frequenz von Zustandsänderungen des abgebildeten Realitätsausschnittes, bedingt durch die Aktualität und den Detailliertheitsgrad operativer Daten.

3.1.2 Analytische Datensicht

Im Unterschied zur operativen ist die analytische Sichtweise durch Betrachtung weitestgehend vergangenheitsbezogener, integrierter und aggregierter Daten charakterisiert. Der Fokus liegt nicht auf der detaillierten Zustandbeschreibung einzelner Objekte, sondern auf der Darstellung von Kennzahlausprägungen in Abhängigkeit von aktuellen und historischen Objekteigenschaften (vgl. Kapitel 2.2.2.1). Das Beispiel der Milchsperrern aus dem letzten Kapitel (3.1.1.1) wieder aufgreifend, interessiert in einer analytischen Sicht nicht die Frage nach einer aktuellen Milchsperrern, sondern beispielsweise nach der durchschnittlichen Anzahl der Tage mit Milchsperrern je Tier in Tiergruppen in Abhängigkeit vom Milchleistungsniveau.

Derartige Erweiterungen des Betrachtungsraumes erfordern ein Schema der Datenverdichtung zur Darstellung von Abhängigkeiten zwischen Objekten. Die Erhöhung der Aussagefähigkeit von Daten erfolgt mittels Aggregation zu Kennzahlen (HIRSCHAUER, 2001, S. 282) bei gleichzeitiger Reduktion der Komplexität des Betrachtungsraumes. Abhängigkeiten zwischen Informationsobjekten, aus denen Kennzahlausprägungen begründet sind, werden als Auswertepfade betrachtet. Die Pfade stellen Strukturen einer Kennzahlaggregation dar. Die entsprechende Semantik wurde in Kapitel 2.2.2.1 unter den Stichworten Multidimensionalität und OLAP-Konzept beschrieben.

Eine Menge vergleichbarer Objekte wird als Objekttyp bezeichnet. Beispielsweise können Tiere als Objekte über einen Objekttyp „Tier“ dargestellt werden. Entsprechend dem bereits eingeführten multidimensionalen Vokabular werden diese auch als Klassifikationsstufen bezeichnet. Ein Klassifikationsschema, das sich aus der Hierarchie von Objekten (multidimensional: Klassifikationshierarchie) vom Tier über die Tiergruppe bis zum Stall ableitet, zeigt Abbildung 8. Informationen über Milchsperrern der Einzeltiere lassen sich damit entlang des dargestellten Klassifikationspfades zu Häufigkeiten von Milchsperrern in Tiergruppen oder Ställen verdichten (Tabelle 1). Bei der Modellierung entsprechender Datensichten sind deshalb nicht mehr nur Datenstrukturen, sondern gleichzeitig die damit verbundene Auswertesemantik in Form von Klassifikationsschemata abzubilden. Dadurch

ergibt sich als weiterer Unterschied zur operativen Datensicht ein klarer Funktionsbezug der analytischen Datensicht, nämlich der einer Analyseunterstützung.

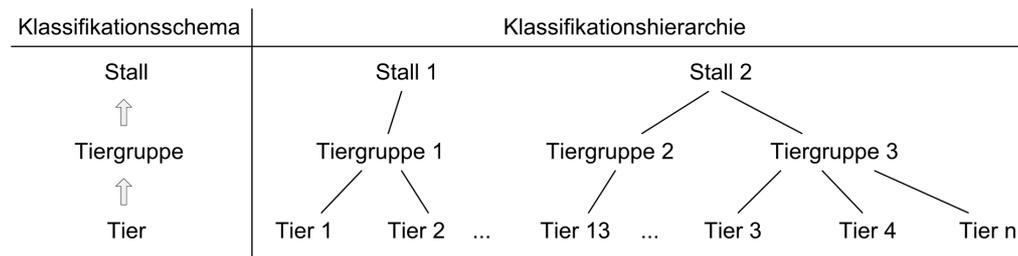


Abbildung 8: Klassifikationsschema und Klassifikationshierarchie

Tabelle 1: Durchschnittliche Anzahl an Tagen mit Milchsperr

	Stall 1			Stall 2						
	Tiergruppe 1		Gesamt	Stall 1 Gesamt	Tiergruppe 2		Tiergruppe 3			Stall 2 Gesamt
	Tier 1	Tier 2			Tier 13	Gesamt	Tier 3	Tier 4	Gesamt	
Jan 06	2,3	0,0	1,2	1,2	12,0	12,0	0,9	23,0	12,0	12,0
Feb 06	2,1	0,0	1,1	1,1	19,0	19,0	4,5	0,0	2,3	7,8
Mrz 06	0,0	1,1	0,6	0,6	9,0	9,0	0,0	3,0	1,5	4,0
I. Quartal 06	1,5	0,4	0,9	0,9	13,3	13,3	1,8	8,7	5,2	7,9

3.2 Modellierungsprozess

Die Modellierung einer Datenbank kann als Prozess mit dem Ziel aufgefasst werden, ausgehend von bestehenden Nutzeranforderungen eine technische Organisation der Datenspeicherung zu entwerfen (Abbildung 9).

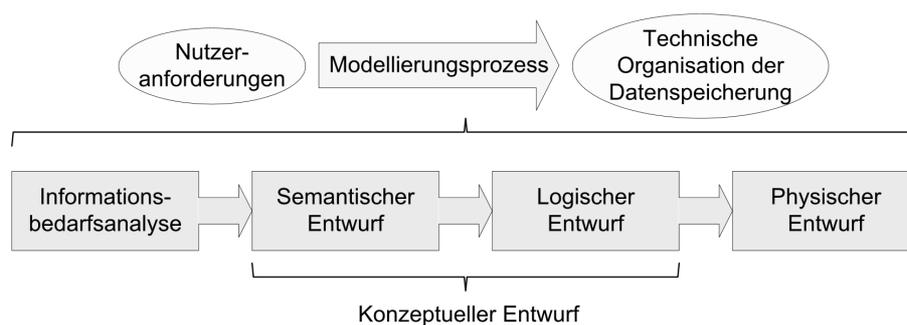


Abbildung 9: Schritte des Datenbankentwurfsprozesses

Nach MAYR et al. (1987, S. 482) sind unter dem Datenbankentwurf „[...] alle Aufgaben und Tätigkeiten zur Ermittlung und Festlegung der für einen Anwendungsfall ‚aktuellen Parameter‘ für Art und Struktur der durch ein gegebenes Datenbanksystem zu verwaltenen Informationen [...]“ zu verstehen. Entsprechend dabei anfallender Aufgaben kann der Modellierungsprozess in eine Reihe aufeinander aufbauender Entwurfsschritte unterteilt

werden, die verschiedene Ebenen der Modellabstraktion darstellen (Abbildung 9). Entgegen der in der Literatur weit verbreiteten Unterscheidung in einen konzeptuellen und einen logischen Entwurf (z.B. ELMASRI und NAVATHE, 2002, S. 571 ff.) wird im Folgenden der logische Entwurf auf Grund seines konzeptuellen Charakters als Teil des konzeptuellen Entwurfes angesehen. Damit wird die von FISCHER (1992, S. 72 f.) vorgenommene Aufspaltung des konzeptuellen Schemas der 3-Schichten-Architektur der ANSI-SPARC in ein semantisches und ein logisches Subschema auf den Modellierungsprozess übertragen.

Ziel der Unterteilung des Entwurfsprozesses in einzelne Entwurfsschritte ist eine Verringerung der Komplexität des Modellierungsprozesses, die zum einen die Handhabbarkeit und Nachvollziehbarkeit der Entwurfsergebnisse für den Nutzer erhöht und zum anderen die Abhängigkeit der Modellierungsergebnisse von der konkreten technischen Realisierung (Hardware, Datenbankmanagementsystem, etc.) reduziert. Besonders bei der Erstellung eines Referenzmodells ist darauf zu achten, den Modellierungsprozess transparent zu gestalten und die entsprechenden Modellierungsschritte einzuhalten. Grundsätzlich ist der Modellierungsprozess als gerichtete Folge der Entwurfsschritte zu sehen, das heißt ein Entwurfsschritt hat unabhängig von der folgenden Entwurfsebene zu sein. Praktisch kann jedoch von einem iterativen Verfahren ausgegangen werden, da Änderungen der abgebildeten Umwelt eingearbeitet oder Fehler vorhergehender Entwurfsschritte korrigiert werden müssen (MAYR et al., 1987, S. 482; KEMPER und EICKLER, 2004, S. 30 f.; ELMASRI und NAVATHE, 2002, S. 571 ff.).

Entsprechend der Entwurfsschritte gliedert sich die weitere Arbeit:

Informationsbedarfsanalyse

Erster Schritt des Modellierungsprozesses ist die Informationsbedarfsanalyse. Sie dient der Beantwortung der Frage, welche Nutzeranforderungen das Datenmodell eines Informationssystems zu erfüllen hat (MAYR et al., 1987, S. 484). Abgeleitet aus den Anforderungsspezifikationen liegen als Ergebnis der Informationsbedarfsanalyse Datenobjekte und deren Eigenschaften vor.

Semantischer Entwurf

Aufgabe des semantischen Entwurfes ist es, die Datenstrukturen der Anforderungsanalyse in ein formales, semantisches Modell zu überführen. Ergebnis ist ein von logischen Datenmodellen unabhängiges Schema der Datenstrukturen des abzubildenden Realitätsausschnittes, das die formalisierte Schnittstelle zum weiteren Modellierungsprozess bildet.

Der semantische Datenbankentwurf wird in der Literatur auch als konzeptueller bzw. konzeptioneller Entwurf bezeichnet. Entsprechend der vorangestellten Argumentation in Bezug auf den logischen und konzeptuellen Entwurf wird statt des konzeptuellen Entwurfes der Begriff des semantischen Entwurfes adäquat zur Bezeichnung semantischer Datenmodelle (FISCHER, 1992, S. 72 f.; LASSMANN et al., 2002, S. 236) verwendet.

Logischer Entwurf

Darauf aufbauend erfolgt die Modellierung der Datenstrukturen mithilfe eines logischen Datenmodells. Hierbei wird das semantische Schema in ein Implementierungsdatenmodell transformiert, wie es letztendlich vom Datenbankmanagementsystem (DBMS) verwendet werden soll. An dieser Stelle unterscheidet man zwischen einem systemunabhängigen Entwurf und einem systemabhängigen Entwurf, bei dem DBMS-spezifische Merkmale enthalten sind.

Physischer Entwurf

Die Anpassung des logischen, DBMS-abhängigen Entwurfs an Hard- und Software sowie das Datenbanknutzungsverhalten wird als physischer Entwurf bezeichnet. Dabei sind unter anderem spezifische Speicherstrukturen und Zugriffspfade festzulegen.

4 Informationsbedarfsanalyse

4.1 Grundlagen

4.1.1 Aufgabe der Informationsbedarfsanalyse

Grundlage jeder Informationssystementwicklung und einer damit verbundenen Datenbankentwicklung ist die Analyse der Anforderungen an das Anwendungssystem. Identifizierte Aufgaben müssen hinreichend informationstechnisch unterstützt und damit im Zusammenhang benötigte Daten in Datenschemata zu erstellender Datenbanken berücksichtigt werden. In dieser Phase begründete Fehler und Ungenauigkeiten lassen sich im späteren Verlauf nur schwer und mit hohem Aufwand berichtigen (STRAUCH und WINTER, 2002, S. 360). Insofern kommt der Bestimmung des Informationsbedarfs eine besondere Bedeutung zu.

Der Informationsbedarf ist definiert als „[...] die Art, Menge und Qualität der Informationen, die eine Person zur Erfüllung ihrer Aufgaben in einer bestimmten Zeit benötigt.“ (PICOT et al., 2001, S. 81). Damit ist der Informationsbedarf stark nutzergeprägt und steht im Spannungsfeld subjektiver Wahrnehmung von Angebot und Nachfrage. Infolgedessen unterscheidet man verschiedene Informationsteilmengen (Abbildung 10). Für weitergehende Erläuterungen sei auf PICOT et al. (2001, S. 79 ff.) und STRAUCH und WINTER (2002, S. 363) verwiesen.

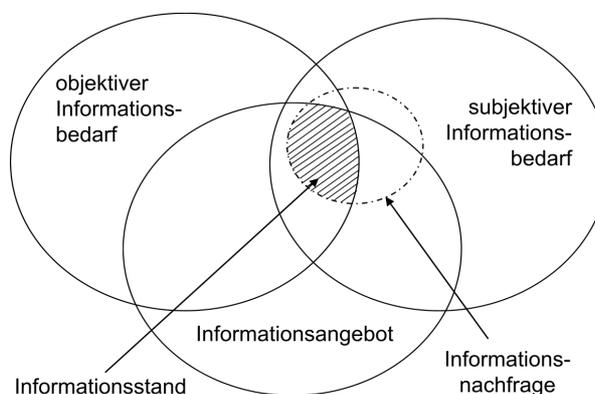


Abbildung 10: Informationsteilmengen (nach PICOT et al., 2001, S. 82)

4.1.2 Analyseansätze zur Ermittlung des der Informationsbedarfs

Die Charakterisierung eines objektiven Informationsbedarfs bei Existenz verschiedener subjektiver Informationsbedarfe erfordert die Anwendung unterschiedlicher Analyseansätze zur Abbildung der verschiedenen Sichtwinkel des Anwendungsfeldes (vgl. STRAUCH und WINTER, 2002, S. 367 f):

▪ Nachfrageorientierter Ansatz

Explizit in der Definition des Informationsbedarfs ist der Nutzer als Konsument von Informationen genannt, dessen Sicht entsprechend durch einen nachfrageorientierten Ansatz der Informationsbedarfsanalyse zu berücksichtigen ist. Hintergrund des Ansatzes ist die Annahme, dass der Nutzer selbst seinen Informationsbedarf am besten einschätzen kann. Nachteile des Ansatzes liegen neben der bereits angeführten Subjektivität, in einer eingeschränkten Fähigkeit des Nutzers, einen zukünftigen und vom aktuellen Realisierungsstand unabhängigen Informationsbedarf formulieren zu können.

▪ Angebotsorientierter Ansatz

In den meisten Anwendungsfällen existieren bereits Informationssysteme, die mit dem aktuellen Projekt in Verbindung stehen. An dieser Stelle setzt ein angebotsorientierter Analyseansatz an, der das Datenangebot bestehender Informationssysteme identifizieren soll. Durch Analyse aller Informationssysteme oder direkt deren Datenschematas im Projektbereich sind Rückschlüsse auf Daten und ihre Strukturen ableitbar, um in der Summe das Informationsangebot zu charakterisieren. Neben einer weitgehenden Nichtbeachtung des Nutzers ist der Hauptkritikpunkt in der Ausblendung zukünftiger Erfordernisse zu sehen.

▪ Geschäftsprozessorientierter Ansatz

Im Rahmen des geschäftsprozessorientierten Ansatzes werden Informationen⁵ aus ablaufenden Geschäftsprozessen betrachtet. Dadurch erhält man einen weitestgehend objektiven Überblick über eine Informationsnutzung. Damit wird die Subjektivität des nachfrageorientierten Ansatzes und die Beschränkung auf das Informationsangebot existierender EDV-unterstützter Informationssysteme des angebotsorientierten Ansatzes umgangen. In Abhängigkeit des Unterstützungsgrades von Geschäftsprozessen durch bestehende EDV-unterstützte Informationssysteme im Anwendungsfeld decken sich jedoch geschäfts- und angebotsorientierter Analyseansatz und teilen infolgedessen auch

⁵ Durch den Sprachgebrauch des *Informations-* und nicht des *Datenbedarfs* wird im Folgenden in diesem Zusammenhang auch von Information gesprochen.

dieselben Nachteile. Ein wesentlicher Informationszuwachs durch den geschäftsprozessorientierten Ansatz ist folglich nur bei einer geringen Abdeckung von Unternehmensaufgaben durch Informationssysteme zu erwarten.

Dessen ungeachtet stellen alle drei Ansätze berechnete Ausgangspunkte für die Analyse des Informationsbedarfs dar.

4.1.3 Methoden der Informationsbedarfsanalyse

Unabhängig der vorgestellten Analyseansätze besitzen die Methoden der Umsetzung entscheidenden Einfluss auf die Qualität der Informationsbedarfsanalyse. Zu nennen sind hier vor allem (vgl. z.B. BEIERSDORF, 1995, S. 71 ff.; VOß und GUTENSCHWAGER, 2001, S. 129 ff.):

- **Das Literaturstudium**

Anhand von Veröffentlichungen lassen sich sowohl Informationsbedürfnisse der Nutzer als auch Informationsangebote identifizieren. Hintergrund der Methode ist die Annahme, dass ein Großteil verfügbaren Wissens in vielfältiger Weise dokumentiert ist.

- **Die Berichts-/Dokumentenanalyse**

Die im Rahmen eines betrieblichen Berichtswesens⁶ transportierten Informationen lassen Rückschlüsse auf angebotene und in Geschäftsprozessen relevante Daten zu. Eine Berichtsanalyse ist deshalb zur Bestimmung des Informationsbedarfs in entsprechenden Ansätzen anwendbar. Das Berichtswesen lässt jedoch nicht zwangsläufig Rückschlüsse auf die Informationsnachfrage zu und ist aus diesem Grund für einen nachfrageorientierten Ansatz weniger geeignet.

- **Die Analyse bestehender Informationssysteme**

Die Analyse bestehender Informationssysteme stellt den klassischen Ansatz zur Ermittlung des Informationsangebotes dar. Daneben sind ebenfalls Informationen zu unterstützten Geschäftsprozessen ableitbar, nicht ohne weiteres jedoch zur Informationsnachfrage.

- **Die Expertenbefragung**

Die Expertenbefragung stellt die Methode der Wahl zur Umsetzung eines nachfrageorientierten Ansatzes dar, ist jedoch für alle anderen Ansätze gleichermaßen geeignet.

⁶ Der Bericht wird an dieser Stelle sehr weit als verdichtete Zusammenstellung von Daten zu Informationszwecken ausgelegt.

Durch die Befragung von Personen mit großem Wissen und Erfahrungen in einem Fachgebiet soll die entsprechende Informationsmenge bestimmt werden. Dabei ist es aus methodischer Sicht unerheblich, ob die Befragung in Interviews oder standardisiert über Fragebögen erfolgt.

Tabelle 2 ordnet die Methoden den Ansätzen der Informationsbedarfsanalyse zu, in denen sie zum Einsatz kommen können.

Tabelle 2: Matrix der Analyseansätze und -methoden für die Informationsbedarfsanalyse

Methoden \ Ansätze	Nachfrageorientiert	Angebotsorientiert	Geschäftsprozessorientiert
Literaturstudium	X	X	X
Berichtsanalyse		X	X
Analyse bestehender Informationssysteme		X	X
Expertenbefragung	X	X	X

4.2 Informationsbedarfsanalyse im Rahmen des Precision Dairy Farming

Das Kapitel charakterisiert die zu berücksichtigenden Aspekte des Informationsbedarfs in der Milcherzeugung und erläutert angewandte Vorgehensweisen und Methoden. Aus der Liste an abzubildenden Merkmalen als Ergebnis der Informationsbedarfsanalyse werden Objekte, Beziehungen und deren Merkmale identifiziert, die sich durch eine Gruppierung nach räumlichen, zeitlichen und sachlichen Gesichtspunkten herausbilden und damit den Ausgangspunkt für den eigentlichen Datenbankentwurf darstellen.

4.2.1 Vorgehensweise und Durchführung

Um ein weitestgehend vollständiges Bild des Informationsbedarfs im Rahmen des Precision Dairy Farming zu erhalten, wurden sowohl der nachfrage-, der angebots- als auch der geschäftsprozessorientierte Analyseansatz verfolgt. Die Anwendung des nachfrageorientierten Ansatzes erschließt sich aus der Aufgabe der Datenmodellierung zur Unterstützung von Nutzeranforderungen. Der angebotsorientierte Ansatz erlangt seine Berechtigung über die starke Prägung des Precision Dairy Farming durch neue technische Möglichkeiten in der Prozessdatenerfassung. Entsprechende Lösungen münden aufgrund ihrer Aktualität zum Teil noch nicht in einer adäquaten Informationsnachfrage, weshalb nur ein angebotsorientierter Ansatz zu Ergebnissen führt. Die Vielzahl und die teilweise noch geringe EDV-Unterstützung interner und externer Geschäftsprozesse landwirtschaft-

licher Unternehmen erfordern dagegen einen geschäftsprozessorientierten Ansatz, um die adäquate Informationsteilmenge zu erfassen.

Die Umsetzung der Analyseansätze erfolgte mit den bereits genannten Methoden Literaturstudium, Berichtsanalyse, Analyse bestehender Informationssysteme und Expertenbefragung. Hauptaugenmerk lag jedoch auf der Expertenbefragung, da sie das wichtigste Werkzeug zur Ermittlung eines zukünftigen Informationsbedarfs darstellt.

Expertenbefragung

Aufgrund der Komplexität des Aufgabengebietes der Milcherzeugung wurden die Expertenbefragungen mithilfe der Interviewtechnik durchgeführt. Die Befragungen wurden als offene Gespräche ohne Fragebogen angelegt. Als Gesprächsgerüst diente eine Systematik der technologischen Teilverfahren der Milcherzeugung, die sowohl die Vollständigkeit der betrachteten Prozesse als auch den Befragten genügend Freiraum für nicht berücksichtigte Aspekte sicherstellen sollte. Betrachtete Teilverfahren sind:

- Milchgewinnung
- Fütterung
- Gesundheits-, Konditions- und Verhaltenskontrolle und -erfassung, Behandlungen
- Reproduktion
- Stallklimaerfassung und -steuerung
- Einstreu und Entmistung
- Dokumentation Bestandesorganisation, Tierkennzeichnung
- Dokumentation Ausstattung, Betriebsdaten

Zur weiteren Untergliederung wurden die Teilverfahren in einzelne Prozesse zerlegt, die eine Konkretisierung der Fragen erlauben (Tabelle A 3). Abbildung 11 zeigt beispielhaft eine Aufstellung der Prozesse für das Teilverfahren der Milchgewinnung.

Das Gesprächsgerüst wurde im Laufe der Informationsbedarfsanalyse dem Wissensstand angepasst, das heißt bis dahin unberücksichtigte Aspekte wurden eingearbeitet und bei folgenden Interviews beachtet. Mit Bezug auf die Aufgabenstellung lag der Schwerpunkt der Analyse auf einer operativen Informationsmenge und den Anforderungen zur Erfüllung analytischer Aufgaben. Es wurden insgesamt 36 Befragungen durchgeführt. Die Interviewpartner vertraten dabei die Fachgebiete Tierhaltung/Ethologie, Tierzucht, Tierernährung, Landtechnik/Verfahrenstechnik, Veterinärmedizin/Tierhygiene, Betriebswirtschaft/Unternehmensführung und IT-Dienstleistung. Den Schwerpunkt bildeten dabei Personen aus den Bereichen Forschung, Beratung und regionale Organisationen.

Milchgewinnung

- Melkung
 - Milchqualitätserfassung
 - Milchabgabe (Mengen und Zeiten)
 - Prozessablauf
 - Eutergeometrie
 - Eutergesundheitserfassung
 - MLP
 - Probenkenndaten
 - Beprobungsergebnisse
 - Tierdaten
 - Sekundärdatenerfassung im Melkstand
 - Aktivitätskennzahlen
 - Physiologische Parameter
 - Milchlagerung
 - Milchablieferung
 - Dokumentation Melktechnik
 - Dokumentation Ausstattung
 - Dokumentation Reinigung
 - Dokumentation Überprüfung von Melk- und Kühltechnik
 - Dokumentation Wartung von Melktechnik
 - Dokumentation technische Störung
-

Abbildung 11: Prozessliste mit Merkmalsgruppen, Ausschnitt Milchgewinnung

4.2.2 Ergebnisse der Informationsbedarfsanalyse

Direktes Ergebnis der Informationsbedarfsanalyse ist eine Liste in der Modellierung zu berücksichtigender Merkmale zur Befriedigung eines operativen Informationsbedarfs und als Basis analyseorientierter Aufgaben. Die Merkmale können aus inhaltlicher Sicht Organisationseinheiten zugeordnet und damit gruppiert werden. Die Organisationseinheiten stellen grundlegende Einheiten in der Datenstrukturierung im Rahmen des Precision Dairy Farming dar:

- Euterviertel
- Einzeltier
- Tiergruppe
- Stallabteil
- Stall
- Abrechnungseinheit
- Betriebsstätte
- Unternehmen

Aus der Zusammenfassung der Merkmale nach Prozess, Ort und Zeitpunkt des Datenanfalls und den Organisationseinheiten der Datenzuordnung resultieren als weiterführendes

Ergebnis sachlich, räumlich und zeitlich definierte Merkmalsgruppen (Tabelle 3, Tabelle A 4). Diese dienten als Ausgangspunkt für die nachfolgende semantische Modellierung. Parallel zur Modellierung wurde die Informationsbedarfsanalyse fortgesetzt und neue Erkenntnisse in den Modellierungsprozess eingearbeitet.

Tabelle 3: Anzahl der Merkmale und Merkmalsgruppen je Teilverfahren und Organisationseinheit (SCHULZE et al., 2004)

		Teilverfahren								Summe
		Milchgewinnung	Fütterung	Gesundheitsdaten, physiologische Parameter	Stallklimaerfassung und -steuerung	Reproduktion	Einstreue und Entmistung	Dokumentation, Kennzeichnung, Bestandesorganisation	Dokumentation Ausstattung, Betriebsdaten	
Organisationseinheit	Euterviertel	1/14*								1/14
	Einzeltier	2/29	2/4	2/11		7/85		3/26		16/155
	Tiergruppe		3/13				2/6	1/3		6/22
	Stalleinheit				2/15		1/8	3/10	1/13	7/46
	Abrechnungseinheit	2/11							1/5	3/16
	Betriebsstätte								1/10	1/10
	ohne eindeutige Zuordnung		3/78	3/29					6/30	12/137
Summe		5/54	8/95	5/40	2/15	7/85	3/14	7/39	9/58	46/400

*Anzahl Merkmalsgruppen / Anzahl Merkmale

4.2.3 Diskussion der Informationsbedarfsanalyse

Die im Ergebnis der Informationsbedarfsanalyse vorliegende Liste an abzubildenden Merkmalen bzw. deren Verdichtung zu Merkmalsgruppen verdeutlicht die Komplexität des Unternehmenszweiges Milcherzeugung. Darin sind 400 Merkmale in 46 Merkmalsgruppen ausgewiesen. Besondere Bedeutung erhält dabei das Einzeltier aufgrund der hohen Anzahl direkt zugeordneter Merkmale, wodurch die einzeltierbezogene Prägung des Precision Dairy Farming bestätigt wird.

Neben den inhaltlichen Ergebnissen können Aussagen zur Methodik der Informationsbedarfsanalyse, insbesondere zur Expertenbefragung getroffen werden. Die Expertenbefragung zeigt sich als besonders geeignet:

- Vorbefragungen zur Strukturierung des Aufgabengebietes durchzuführen,
- Merkmalskomplexe mit zukünftiger Bedeutung zu bestimmen und

- Detailfragen in abgegrenzten Anwendungsbereichen zu klären.

Zur Erfassung eines detaillierten, umfassenden operativen Informationsbedarfs erscheint die Expertenbefragung dagegen weniger geeignet. Hierfür sind andere, vornehmlich dokumentenbasierte Methoden vorzuziehen. Für die Analyse des Informationsbedarfs eines komplexen Aufgabengebietes sind daher verschiedene Methoden zu kombinieren.

5 Semantischer Entwurf

Die Aufgabe des semantischen Modellierungsschrittes ist die formale Beschreibung des Realitätsausschnittes, der durch die Informationsbedarfsanalyse definiert und dokumentiert ist. Die Modellierungssprache muss deshalb in der Lage sein, die Semantik des Abbildungsausschnittes in ein Schema zu überführen.

Neben den anwendungsspezifischen Modellierungsanforderungen werden nachfolgend die verwendeten Modellierungsnotationen (E/RM und mE/RM) und angewandten Modellierungsgrundsätze erläutert. Die Modellierung analytischer Daten erfordert die Einführung neuer Notationselemente, um den Anwendungskontext in seiner Vollständigkeit abzubilden. Zusätzlich sind Aggregationstypen von Kennzahlen zu diskutieren. Anhand ausgewählter Beispiele werden nachfolgend semantische Schemata für operative Daten und deren Erweiterung zur Abbildung einer analyseorientierten Sichtweise vorgestellt.

5.1 Überblick über existierende Modellierungsansätze

5.1.1 Modellierungstechniken für operative Daten

Die operative Datensicht ist weitestgehend semantikarm und beschränkt sich auf die grundlegende Unterscheidung von

- [a] Objekten,
- [b] Beziehungen zwischen Objekten und
- [c] Eigenschaften von Objekten und Beziehungen (Kapitel 3.1.1.1),

die in entsprechenden Modellierungsansätzen Berücksichtigung finden müssen. Nachfolgend werden einige Ansätze im Überblick vorgestellt, die diese Anforderungen erfüllen und im Rahmen der operativen Datenmodellierung gebräuchlich sind.

Entity-Relationship Modell (E/RM) und seine Erweiterungen

Der wohl verbreitetste Modellierungsansatz für operative Informationssysteme ist der Entity-Relationship-Ansatz nach CHEN (1976). Definierte Elemente des Modells sind:

- Entitätstypen zur Abbildung gleichartiger Objekte als Objekttyp,
- Beziehungstypen zur Abbildung gleichartiger Beziehungen zwischen Entitäten,

- Attribute und
- Primärschlüssel zur Identifizierung von Entitäten in Entitätstypen.

Zusätzlich erfolgt eine Unterscheidung von Entitäts- und Beziehungstypen in schwache und reguläre Typen. Schwache Entitätstypen werden von einem anderen Entitätstyp bestimmt und werden deshalb auch über einen schwachen Beziehungstyp identifiziert (vgl. Kapitel 5.2.1.2). Als wichtige Erweiterung des E/RM ist die Einbindung von Konzepten der Spezialisierung/Generalisierung zur Gruppierung gleichartiger Entitätstypen zu nennen. Enthalten sind diese Konzepte im **erweiterten Entity-Relationship-Modell (EE/RM)** (ELMASRI und NAVATHE, 2002, S. 97 ff.). Die Erweiterung besteht in der Einführung eines übergeordneten Entitätstyps für Entitätstypen mit einer Anzahl gleicher Merkmale (Generalisierung) bzw. im umgekehrten Fall, der Einführung von Sub-Entitäten, die sich in bestimmten Merkmalen unterscheiden (Spezialisierung). Der Entitätstyp „Maschine“ ist beispielsweise durch die Attribute „Hersteller“ und „Baujahr“ gekennzeichnet. Spezialisierungen der „Maschine“ stellen z.B. die Entitätstypen „Melkstand“ und „Futtermischwagen“ dar. Beide werden durch die gleichen Attribute „Hersteller“ und „Baujahr“ beschrieben, unterscheiden sich aber beispielsweise durch die Attribute „Anzahl Melkplätze“ und „Volumen Mischbehälter“.

Eine Modifikation des E/RM hinsichtlich der Struktur von Beziehungen liegt mit dem **strukturierten Entity-Relationship Modell (SERM)** vor. Der universelle Beziehungstyp wird um einen Entitäts-Beziehungs-Typ erweitert, der existenzielle Abhängigkeiten zwischen Entitätstypen abbildet. Dadurch werden semantische Inkonsistenzen durch Eliminierung von Zirkelbezügen vermieden, Schemata durch die Reduzierung von Beziehungstypen kleiner und die grafische Repräsentation der Schemata durch die Modellierung von Existenzabhängigkeiten und quasi-hierarchische Anordnung der Elemente lesbarer (BOEHNLEIN und ULBRICH-VOM ENDE, 1999).

Unified Modeling Language (UML)

Obwohl das ursprüngliche Anwendungsgebiet der UML der objektorientierte Softwareentwurf ist, wird sie auch zur Datenmodellierung verwendet. Die Terminologie, der für die statische Datenmodellierung verwendeten UML-Elemente, ähnelt der des EE/RM:

Entitätstypen werden als Klassen und Entitäten als Objekte bezeichnet. Ebenso haben Beziehungstypen ihre Entsprechung in Assoziationen und Beziehungsinstanzen in Verknüpfungen. Weiterhin existieren die Konzepte des Attributes und der Vererbung (BALZERT, 2005, S. 8 ff., 19 ff.; OMG, 2005, S. 48 ff., 67 ff.).

5.1.2 Modellierungstechniken für analytische Daten

Bedingt durch den Funktionsbezug einer analytischen Datensicht (Kapitel 3.1.2) erfordert die Abbildung analytischer Daten zusätzliche Modellsemantik im Vergleich zu operativ ausgerichteten Modellierungsansätzen. Diese besteht in der Unterscheidung von qualifizierenden und quantifizierenden Daten als Grundgedanke einer multidimensionalen Sichtweise. Es muss deshalb grundsätzlich

- [a] die Identifizierung von quantifizierenden Datenstrukturen,
- [b] die Identifizierung von qualifizierenden Datenstrukturen und
- [c] die Abbildung von Beziehungen zwischen qualifizierenden Objekten und der daraus folgenden Ausbildung von Auswertepfaden erfolgen.

Mit den bereits vorgestellten Modellierungsansätzen ist eine adäquate Abbildung nur unzureichend möglich, da eine Unterscheidung quantifizierender und qualifizierender Datenstrukturen fehlt. Die multidimensionale Semantik ist in entsprechenden Schemata nur indirekt ableitbar (GOLFARELLI et al., 1998a; BOEHNLEIN und ULBRICH-VOM ENDE, 1999). So kann beispielsweise aus den Kardinalitäten im E/RM indirekt auf den Beziehungstyp (quantifizierend oder qualifizierend) geschlossen werden. Im Sinne praktikabler Modellierungsansätze sind derartige Lösungen jedoch nicht zu verstehen. Entsprechendes gilt auch für das EE/RM, das SERM⁷ und die UML, weshalb eine Reihe spezieller multidimensionaler Modellierungsansätze entwickelt wurde.

Multidimensionales Entity-Relationship Modell (mE/RM)

Das mE/RM ist zu den evolutionären Modellierungsansätzen zu zählen. Das heißt, es handelt sich um eine Erweiterung/Weiterentwicklung eines bestehenden Modellierungsansatzes, in diesem Fall des bekannten E/RM. Mit dem mE/RM werden drei Ziele verfolgt: Minimalität (die Zahl der Modell-Erweiterungen/Modifikationen ist auf ein Minimum zu beschränken); Spezialisierung (Modellerweiterungen sind als Spezialfälle bestehender Modellelemente zu begreifen) und Vollständigkeit (grundlegende multidimensionale Semantik ist vollständig durch die Modellerweiterung abzubilden) (SAPIA et al., 1998; BAUER und GÜNZEL, 2004, S. 165 ff.). Aus diesen Gründen wurden ein spezieller Entitätstyp zur Abbildung von Klassifikationsstufen und zwei spezielle Beziehungstypen zur Darstellung von Klassifikationsbeziehungen und Faktbeziehungen eingeführt. Im Grundsatz der Minimalität des mE/RM liegen sowohl der Hauptvorteil als auch der

⁷ Vorteil des SERM ist jedoch der Entity-Relationship-Typ als semantische Entsprechung einer Klassifikationsbeziehung.

Hauptnachteil begründet: intuitiv lesbare Schemata, die jedoch nur grundlegende multidimensionale Semantik abbilden können.

Dimensional Fact Modell (DFM)

Beim Dimensional Fact Model handelt es sich um einen grafischen multidimensionalen Modellierungsansatz. Neben den grundlegenden Elementen Dimension, Fakt(-Beziehung) und Kennzahl liegt die Besonderheit in der Unterscheidung zwischen dimensionalen und nicht-dimensionalen Attributen und der expliziten Darstellung von Aggregationsoperationen im Modellschema (GOLFARELLI et al., 1998b; SCHELP, 1999). Hierarchien werden als Verkettung von Knoten (Dimensionsstufen) in Baumstrukturen dargestellt. Bestandteil des Ansatzes sind weiterhin Abfragemuster (Kennzeichnung von Knoten in vordefinierten Abfragen) und eine Methodik zur Ableitung eines Dimensional Fact Schemas aus ER-Schemata (GOLFARELLI et al., 1998a und 1998b).

ADAPT (Application Design for Analytical Processing Technologies)

Ein sehr umfassender Ansatz für die konzeptuelle Modellierung analytischer Daten liegt mit ADAPT vor (GABRIEL und GLUCHOWSKI, 1998; SCHELP, 1999; BULOS und FORSMAN, 2006). Zentrale Elemente sind der Würfel (Hypercube) und Dimensionen. Darüber hinaus existieren weitere Elemente zur Abbildung von Hierarchien, Hierarchiestufen, Dimensionselementen, Dimensionsattributen und Berechnungsformeln. Ferner werden 6 verschiedene Dimensionstypen und 3 Beziehungstypen unterschieden. Den Rahmen der semantischen Datenmodellierung verlässt der Ansatz durch Elemente zur Definition logischer und physischer Strukturen wie z.B. Datenquellen und zur Abbildung von Benutzersystemen (Tabellenkalkulation etc.). Die sehr umfangreiche Palette der Darstellungselemente erschwert jedoch eine intuitive Lesbarkeit und wird als bisher ungenügend definiert kritisiert (GABRIEL und GLUCHOWSKI, 1998; SCHELP, 1999).

Multidimensional Unified Modeling Language (mUML)

HARREN und HERDEN (1999) stellen die mUML als grafischen Ansatz zur Umsetzung von Entwürfen in der Multidimensional Modeling Language (MML) vor. Unter Verwendung von Stereotypen im Sinne von speziellen Klassen für einen Anwendungskontext wurde die Semantik von UML-Elementen multidimensional erweitert, um MML-Schemata über herkömmliche CASE⁸-Werkzeuge modellieren zu können.

⁸ CASE – Computer Aided Software Engineering

5.1.3 Aspekte der Zeitmodellierung im semantischen Entwurf

Neben der Wahl des Modells als Grundlage des semantischen Entwurfes kommt der Darstellung der Zeit eine große Bedeutung zu.

Die Analyse der Ausgangssituation zeigt für den vorliegenden Anwendungsfall eine hohe Zeitabhängigkeit der Daten und gleichzeitig die besondere Relevanz der Zeit in den Konzepten der Datenorganisation. Aus diesem Grund erscheint es notwendig, die Zeitbehandlung im Rahmen der semantischen Datenmodellierung in einem notwendigen Maße zu beleuchten. Auch an dieser Stelle ist eine getrennte Betrachtung operativer und analytischer Daten angebracht.

5.1.3.1 Aspekte der Zeitmodellierung im operativen Entwurf

Klassische operative Datenbanken repräsentieren den aktuellen Zustand eines Systems. Sie werden deshalb auch als „Schnappschuss-Datenbanken“ bezeichnet. Sobald die Abbildung vergangenheitsbezogener Daten gefordert wird, muss eine Versionisierung der Daten erfolgen, das heißt es existieren mehrere zeitbezogene Versionen des gesamten Realitätsausschnittes, von Objekten, Beziehungen oder Attributen (KNOLMAYER und MYRACH, 1996). Die Versionisierung wird im Folgenden auch als Zeitstempelung bezeichnet, da jede zeitbezogene Version eines Elementes durch einen Zeitstempel identifiziert werden muss, um die zeitliche Gültigkeit zu charakterisieren. Die Beispiele werden anhand von Entity-Relationship-Modell-Diagrammen erläutert und damit die Einführung entsprechender Darstellungselemente in Kapitel 5.2.1.2 vorweggenommen.

Grundsätzlich von Belang sind drei Zeitarten: die Gültigkeitszeit, die Transaktionszeit und die benutzerdefinierte Zeit (JENSEN und DYRESON, 1998). Sie beschreiben die Semantik eines Zeitattributes. Gültigkeitszeiten geben das Zeitintervall an, in dem ein Zustand im Realitätsausschnitt Gültigkeit besitzt, das heißt es handelt sich um eine reale Gültigkeit. Transaktionszeiten markieren dagegen den Zeitpunkt, zu dem eine Änderung in der Datenbank registriert wird, unabhängig vom Zeitpunkt der Änderung im Realitätsausschnitt. Sie bilden folglich nur eine Systemgültigkeit ab. Benutzerdefinierte Zeiten stellen Zeiten mit einer vom Nutzer definierten Bedeutung dar, beispielsweise ein Geburtsdatum. In nicht temporal erweiterten Datenbanksystemen dienen sie zur Umsetzung von Gültigkeits- und Transaktionszeiten.

Zur Abbildung von Gültigkeitszeiten im semantischen Entwurf kommen prinzipiell zwei Varianten in Frage:

- [a] die Erweiterung von Objekt- und Beziehungstypen um Zeitattribute und
- [b] die Modellierung der Zeit als eigenständigen Objekttyp.

Die aus beiden Varianten resultierenden semantischen Probleme werden von MYRACH (2005, 322 ff.) ausführlich diskutiert.

Pragmatisch bedeuten beide Varianten die Aufspaltung von Objekttypen in Attributmengen mit gleichem zeitlichem Änderungshorizont, um Redundanz zu verhindern. Als Beispiel sollen Personaldaten dienen (Abbildung 12): Der Objekttyp „Person“ besitzt neben dem Primärschlüsselattribut „Personalnummer“ die Attribute „Geburtsdatum“, „Geburtsort“ und „Familienstand“. Zeitliche Änderungen von Daten sind verständlicherweise nur beim Attribut „Familienstand“ zu erwarten. Bezieht sich nun die Zeitstempelung durch Hinzufügen einer zeitlichen Gültigkeit auf den ganzen Entitätstyp, z.B. durch ein Attribut „gültig_ab“, so gilt die Zeitstempelung auch für die Attribute „Geburtsdatum“ und „Geburtsort“. Eine neue Version einer Entität aufgrund einer Änderung des Familienstandes würde somit auch diese beiden (unveränderten) Attributwerte enthalten. Durch Abspalten des Attributes „Familienstand“ in einen neuen, zeitgestempelten, schwachen Entitätstyp „Familienstand“ wird die Redundanzquelle eliminiert (Abbildung 12).

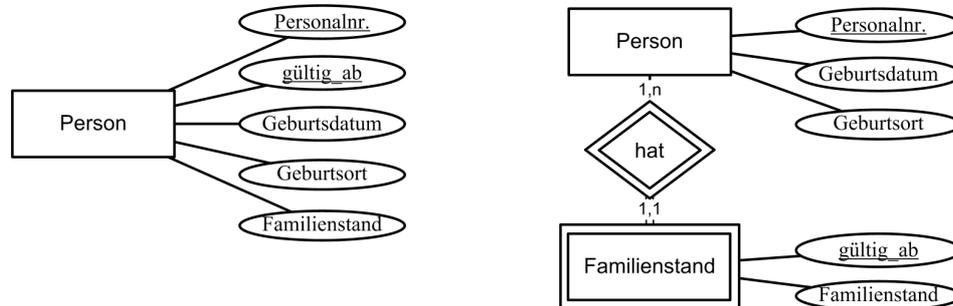


Abbildung 12: Beispiel einer Attributversionisierung durch Entitätstypspaltung im E/RM

Bei der Modellierung eines Zeit-Entitätstyps trifft gleiches zu. Ein Unterschied ergibt sich jedoch hinsichtlich der Zeitstempelung von Beziehungstypen. Die Definition von Schlüsselattributen in Beziehungstypen ist nicht möglich, so dass eine Versionisierung über ein Zeit-Attribut nicht möglich ist. Anders verhält es sich bei Modellierung mittels eines Zeit-Entitätstyps, durch den der Zeitstempel Eingang in die Identifikation des Beziehungstyps erhält (Abbildung 13). Aus dem binären Beziehungstyp zwischen „Person“ und „Unternehmen“ wird dadurch ein ternärer Beziehungstyp. Zu beachten ist jedoch die Ausbildung der Identifikation des Beziehungstyps über alle beteiligten Entitätstypen, die nicht immer korrekt den Sachverhalt abbildet. Kann beispielsweise eine Person zu einer Zeit vertraglich nur in einem Unternehmen gebunden sein, ist die Identifikation der Beziehung über den

Entitätstyp „Unternehmen“ überflüssig, da „Person“ und „Zeit“ die Eindeutigkeit sicherstellen.

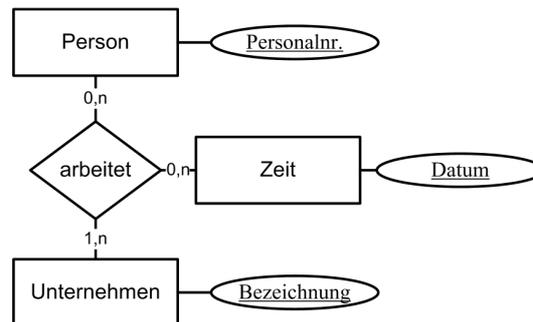


Abbildung 13: Versionisierung eines Beziehungstyps mittels Zeit-Entitätstyp im E/RM

Beide Beispiele verdeutlichen, dass die Zeitstempelung bei Nutzung konventioneller Modellierungsansätze möglich ist, jedoch eine Erhöhung der Modellkomplexität nach sich zieht. Aus diesem Grund existieren verschiedene zeitbezogene Modellierungsansätze, die durch spezielle temporale Modellierungselemente den semantischen Entwurf von Zeitabhängigkeiten vereinfachen (KNOLMAYER und MYRACH, 1996; KAISER, 2000, S. 159 ff.; GÜNZEL, 2001, S. 29 ff.; MYRACH, 2005, S. 333 ff.). Aufgrund der geringen Verbreitung entsprechender Lösungen sollen sie im Folgenden keine Beachtung finden.

5.1.3.2 Aspekte der Zeitmodellierung im analytischen Entwurf

Im Rahmen eines analytischen semantischen Entwurfes muss die Diskussion der Zeit auf Ebene quantitativer (Datenwürfel und Kennzahlen) und qualitativer Strukturen (Dimensionen) geführt werden. Die multidimensionale Datenmodellierung im Rahmen von Data Warehouse- und OLAP-Konzepten impliziert zwangsläufig die Modellierung historischer Daten. Ausdruck dessen ist die Modellierung von Zeit-Dimensionen und der damit verbundenen Zeitabhängigkeit von Kennzahlen. Doch nicht nur Kennzahlen unterliegen zeitlichen Veränderungen, sondern auch die oftmals als statisch betrachteten Dimensionen (GÜNZEL, 2001, S. 37 ff.). Zum einen sind Änderungen in Klassifikationshierarchien möglich, zum anderen sind selbst die Klassifikationsschemata nicht statisch. Der Unterschied wird in Abbildung 14 verdeutlicht. Das Klassifikationsschema, das ein Tier einer Tiergruppe und diese wiederum einem Stall zuordnet (Abbildung 14, links) kann sich ändern. Beispielsweise ist es denkbar, dass die Tiergruppe als Organisationseinheit aufgrund einer vollständigen Einzeltierbewirtschaftung zukünftig unnötig wird. Die Gültigkeit der Klassifikationsstufe und damit die Gültigkeit seiner funktionalen Abhängigkeiten verändern sich. Daraus resultiert eine Änderung eines Pfades im Klassifikationsschema, die entsprechend im semantischen Schema Berücksichtigung finden muss. Noch weit häufiger

kann sich die Klassifikationshierarchie ändern. Im Beispiel betrifft das den Wechsel eines Tieres zwischen Tiergruppen (Abbildung 14, rechts). Dabei ändert sich die Einordnung des Tieres in der Hierarchie der Objekte. Die Zuordnung der Klassifikationsstufe „Tier“ zur Klassifikationsstufe „Tiergruppe“ bleibt jedoch unverändert. Ein semantischer Entwurfsansatz muss deshalb zwischen der Dimensions- und der Instanzänderung unterscheiden.

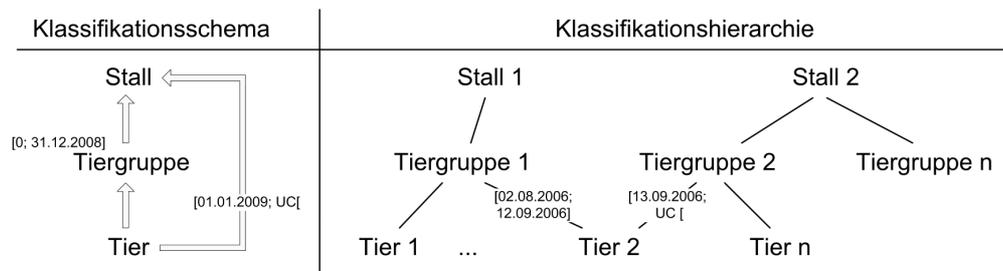


Abbildung 14: Veränderungen in Klassifikationsschemata und Klassifikationshierarchien
 [Intervallanfang; Intervallende] – geschlossenes Intervall;
 [Intervallanfang; UC[- nach oben offenes Intervall (UC – until change)

5.2 Semantischer Datenbankentwurf im Rahmen des Precision Dairy Farming

5.2.1 Semantische Modellierung operativer Daten

5.2.1.1 Anwendungsspezifische Modellierungsanforderungen

Die Anforderungen an die Datenmodellierung im Rahmen des Precision Dairy Farming wurden in Kapitel 2.1.2 dargestellt. Die Modellierung operativer Daten betreffen dabei insbesondere die Forderung nach einem integrierten Datenschema im Sinne einer Datenintegration der verschiedenen operativen Anwendungsbereiche und die Abbildung zeitbezogener Daten zur Erfüllung von Dokumentationsaufgaben.

5.2.1.2 Modellierungsnotation

Als Modellierungsansatz wurde der Entity-Relationship Ansatz gewählt. Er zählt zu den verbreitetsten Ansätzen für den Entwurf operativer Datenschemata und seine weitestgehend leichte Les- und Erlernbarkeit prädestinieren ihn für die Referenzmodellierung, um den Nutzerkreis durch die Wahl des Modellierungsansatzes nicht unnötig einzuschränken. Daneben werden Teile des erweiterten E/RM angewandt, um die Konzepte der Spezialisierung/Generalisierung nutzen zu können.

Die verwendeten Modellierungsobjekte Entitäts- und schwacher Entitätstyp, Beziehungs- und identifizierender Beziehungstyp, Kardinalitäten in min/max-Notation, (Primärschlüssel-)Attribut und Spezialisierung sind in Abbildung 15 und Abbildung 16 dargestellt. Kardinalitäten beschreiben die Komplexität von Beziehungen zwischen Entitäten, das heißt sie geben an, mit wie vielen Entitäten eines anderen Entitätstyps eine Entität in Beziehung stehen muss oder kann. Die Kardinalitäten in Abbildung 15 sind folgendermaßen zu lesen: Eine Entität des Entitätstyps 1 steht mindestens mit 0, maximal mit n Entitäten des Entitätstyps 1 in Beziehung. Die Bezeichnung eines Beziehungstyps als 1:n- oder n:m-Beziehungstyp verwendet als Kurzform nur die maximale Anzahl an Ausprägungen eines Beziehungstyps. So stellt der Beziehungstyp in Abbildung 15 einen n:m-Beziehungstyp und der in Abbildung 16 einen 1:n-Beziehungstyp dar.

Die in Abbildung 16 verwendete Spezialisierung ist total und disjunkt. Total bezeichnet dabei den Umstand, dass alle Entitäten des Super-Entitätstyps Mitglied eines der Sub-Entitätstypen sind. Das heißt, es existiert keine Entität des Super-Entitätstyps, die nicht Entität von Sub-Entitätstyp 1 oder Sub-Entitätstyp 2 ist. Die Charakterisierung als disjunkt gibt an, dass die Sub-Entitätstypen der Spezialisierung nicht überlappend sind. Es gibt also keine Entität des Super-Entitätstyps, die in mehr als einem Sub-Entitätstyp Mitglied ist. Daneben existieren partielle und überlappende Spezialisierungen, auf die an dieser Stelle jedoch nicht weiter eingegangen werden soll (z.B. ELMASRI und NAVATHE, 2002, S. 105 f.).

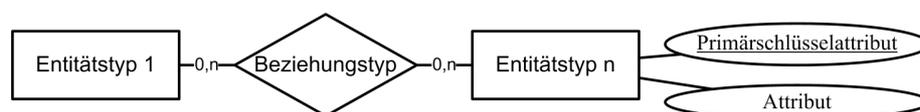


Abbildung 15: E/RM-Basiselemente: Entitätstyp, Beziehungstyp, (Primärschlüssel-) Attribut und Kardinalitäten in min-max-Notation

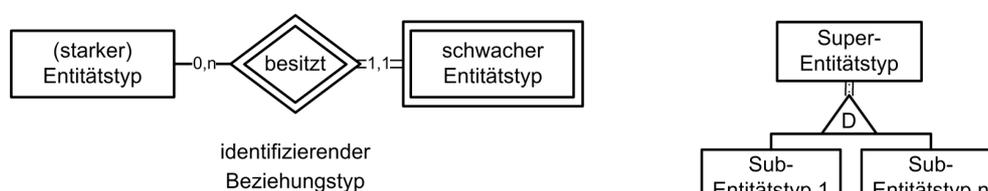


Abbildung 16: Zusätzliche E/RM-Konstrukte: Schwacher Entitätstyp mit identifizierendem Beziehungstyp (links) und totale, disjunkte Generalisierung/Spezialisierung (rechts)

5.2.1.3 Modellierungsgrundsätze

Die Modellierung eines komplexen integrierten Datenschemas erfordert die Festlegung schemaweiter Modellierungsgrundsätze, die nachfolgend genannt werden sollen.

Schemaweiter Zeit-Entitätstyp

Einer der wichtigsten Grundsätze liegt in der Repräsentation einer schemaweit identischen Zeitdefinition durch einen schemaweiten Zeit-Entitätstyp. Somit ist sichergestellt, dass alle Zeit-Entitäten der gleichen Zeitzone angehören, der gleichen Formatierung unterliegen, aber auch in der gleichen, notwendigerweise der niedrigsten Granularität vorliegen. Letzteres stellt insofern ein Problem dar, dass beispielsweise ein Geburtsdatum semantisch nicht korrekt über einen SQL-Datentyp „Timestamp“ als Datum + Uhrzeit abgebildet wird. Der Datentyp impliziert eine zusätzliche Information (die der Uhrzeit), die nicht Teil der Attributausprägung ist. Die Modellierungsvariante bedeutet jedoch keinen Informationsverlust, so dass der semantische Nachteil in Kauf genommen wird.

Gültigkeitszeiten

Gültigkeitszeiten von Entitäten und Attributen werden mittels Anfangs- und Endzeitpunkten modelliert. Unter der Maßgabe der Verwendung eines Zeit-Entitätstyps bedeutet das die Modellierung zweier Beziehungstypen zwischen einem gültigkeitsbeschränktem Entitätstyp und dem Zeit-Entitätstyp. Vorteil dieser Variante ist die semantisch korrekte Darstellung einer Gültigkeit mit Anfangs- und Endpunkt. Die Alternative wäre eine Modellierung nur des Anfangspunktes der Gültigkeit, so dass sich der Endpunkt vom Anfangspunkt der chronologisch folgenden Entitätsversion ergeben würde. Diese Variante impliziert jedoch zwangsläufig eine nicht grundsätzlich gegebene lückenlose Abfolge von Versionen (Gültigkeiten) einer Entität. Für grundsätzliche Fragen zu Zeitintervallen und temporaler Logik wird z.B. auf ALLEN (1983) verwiesen.

Typ-Entitätstypen

Weiterhin wird die Nutzung von Typ-Entitätstypen⁹ favorisiert. Typ-Entitätstypen stellen eine Transformation von Attributausprägungen zu Entitäten dar. Dadurch ist es möglich, mehrfach verwendete, identische Merkmale global über einen Entitätstyp abzubilden und somit die schemaweit identische Merkmalsverwendung sicherzustellen (Abbildung 17). Letztlich stellt der soeben eingeführte Zeit-Entitätstyp ebenfalls einen Typ-Entitätstyp dar. Ebenso Anwendung finden Typ-Entitätstypen bei Merkmalen für die standardisierte Merkmalschlüssel existieren, beispielsweise für den ADR-Rinderrassenschlüssel.

⁹ synonym: Katalog-Entitätstypen

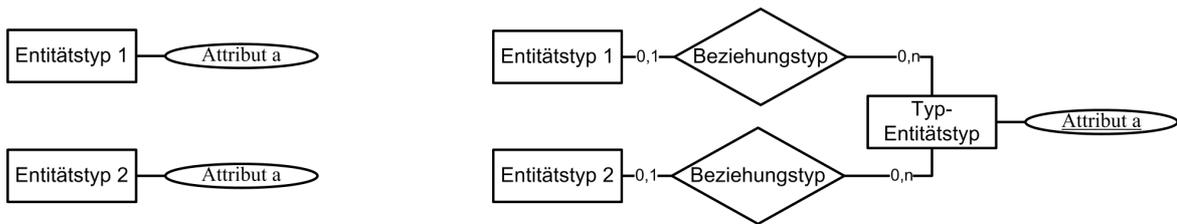


Abbildung 17: Verwendung von Typ-Entitätstypen zur Zusammenführung identischer Attribute

Strukturanalogien

Aufgrund der im vorliegenden Anwendungsfall zu erwartenden Komplexität des Datenschemas ist auf die Identifizierung von Strukturanalogien zur Komplexitätsreduktion Wert zu legen (FETTKE und LOOS, 2005). Dabei ist ein Kompromiss zwischen einer stark generalisierten, jedoch semantisch weitestgehend unspezifischen Modellstruktur und einer fachlich spezifischen, dafür aber komplexen Modellierung zu finden.

Ereignismodellierung

Bei Nutzung von Zeit- und Typ-Entitätstypen stellen Ereignisse im E/RM grundsätzlich Beziehungstypen dar (Abbildung 18). Die Identifizierung eines Ereignisses erfolgt über die Primärschlüsselattribute der beteiligten Entitätstypen.

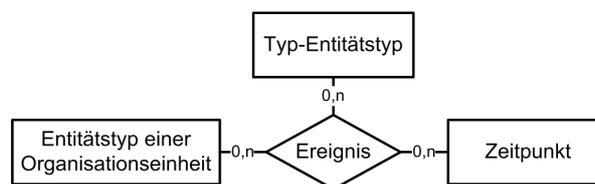


Abbildung 18: Ereignis als Beziehungstyp

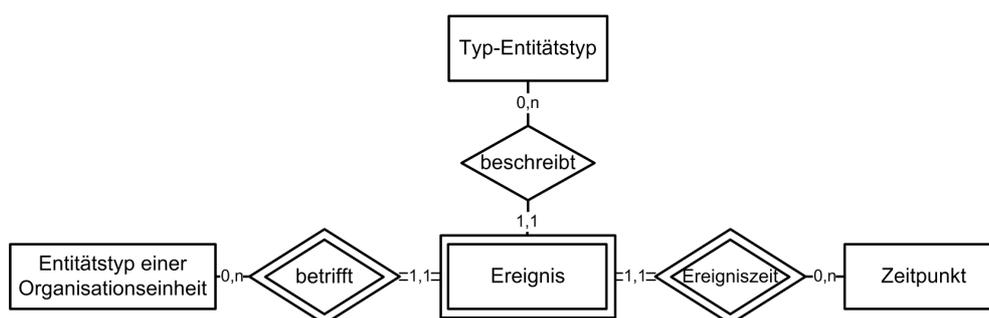


Abbildung 19: Ereignis als schwacher Entitätstyp

Ein Ereignis wie in Abbildung 18 abgebildet, hätte demnach eine dreiteilige Identifikation über die beteiligten Entitätstypen. Hat der Typ-Entitätstyp jedoch nur die Bedeutung einer standardisierten Merkmal-Abbildung, ist die Aufnahme seines Primärschlüssels zur Identifikation unnötig. Entsprechend muss der Ereignis-Beziehungstyp in einen mehrwertig

schwachen Entitätstyp überführt werden, um eine korrekte Schlüsselausprägung zu gewährleisten (Abbildung 19).

Primärschlüssel - Identifikationsnummer

Die Nutzung einer künstlichen Identifikationsnummer (ID) von Entitäten verfolgt im vorliegenden Projekt hauptsächlich den Zweck der Verkleinerung des Primärschlüssels von Entitätstypen. Bilden Entitätstypen mit zusammengesetztem Primärschlüssel (beispielsweise der Ereignis-Entitätstyp in Abbildung 19) n:m-Beziehungstypen aus oder werden selbst zum identifizierenden Entitätstyp eines schwachen Entitätstyps, so entstehen Kaskaden von Primärschlüsselattributen, die eine Lesbarkeit entsprechender Schemata einschränken (Abbildung 20). Aus diesem Grund wurde der Grundsatz verfolgt, referenzierte Entitätstypen mit zusammengesetztem Primärschlüssel mit einer künstlichen Identifikationsnummer zu versehen, um derartige Schlüsselkaskaden zu unterbinden. Die Nutzung einer Primärschlüssel-ID widerspricht jedoch dem Konzept des schwachen Entitätstyps, so dass die Einführung einer ID einen schwachen Entitätstyp zu einem regulären Entitätstyp aufwertet (Abbildung 21).

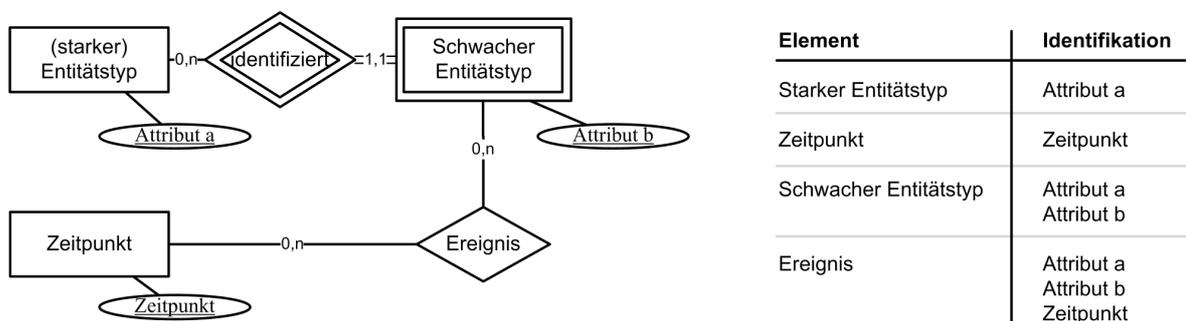


Abbildung 20: Ausbildung von Schlüsselkaskaden

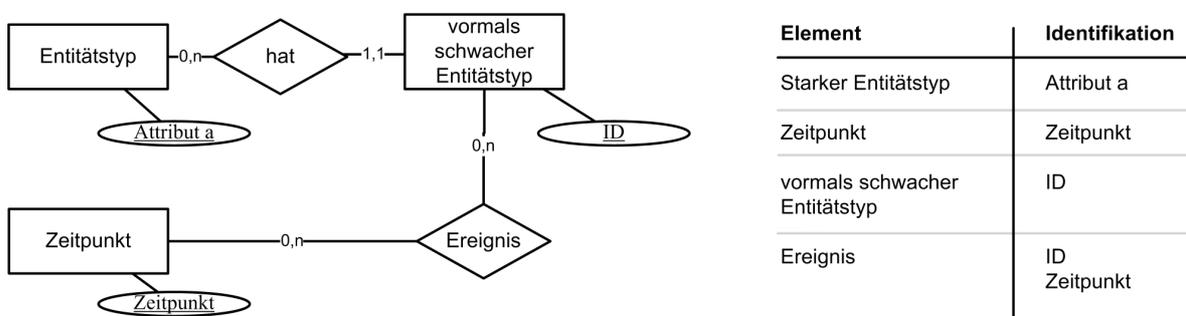


Abbildung 21: Nutzung einer ID zur Verhinderung von Schlüsselkaskaden

5.2.1.4 Ausgewählte Ergebnisse

Aufgrund der Komplexität des semantischen Datenschemas für den Anwendungsbereich des Precision Dairy Farming werden im Folgenden nur einige ausgewählte Ergebnisse der Modellierung vorgestellt. Gegenstand der Ausführungen werden das Tier und seine nachgelagerten Strukturen in der Produktionsorganisation sowie die Modellierung von Krankheitsdaten als wichtiger, bisher wenig beachteter Bereich der Datenmodellierung sein. Daneben wird anhand der Teilmodelle für Fütterungsdaten und für technische Objekte das Vorgehen zur Beseitigung von Strukturanalogien besprochen.

Modellierung des Tieres

Aus der Informationsbedarfsanalyse geht das Tier als wichtigstes Objekt im Precision Dairy Farming hervor. Ein entsprechend hoher Stellenwert kommt ihm deshalb bei der Modellierung im E/RM zu. Das Tier wird als Entitätstyp abgebildet, wobei mittels Spezialisierung eine Unterscheidung in männliche und weibliche Tiere vorgenommen wird (Abbildung 22).

Vererbt werden dabei unter anderem die Lebensohrmarke (LOM) als Primärschlüssel, das Geburtsdatum als Beziehungstyp mit dem Zeit-Entitätstyp und die Angabe von Vater und Mutter als binäre Beziehungstypen auf den Entitätstyp „Tier“ selbst, da sie ebenfalls Entitäten des „Tier“-Entitätstyps darstellen. Von Interesse im Anwendungsgebiet ist weiterhin das Euterviertel als Körperteil eines weiblichen Tieres. Die damit bestehende Existenzabhängigkeit entspricht der Bedeutung eines schwachen Entitätstyps. Aus Gründen einer übersichtlicheren Referenzierbarkeit wird dennoch eine ID als Surrogatschlüssel eingeführt, wodurch der schwache Entitätstyp aufgelöst ist, dessen Bedeutung jedoch über den obligaten Beziehungstyp zum Entitätstyp „Tier“ erhalten bleibt. Über einen Beziehungstyp mit dem Zeit-Entitätstyp besitzt das „Euterviertel“ eine Gültigkeit zur Abbildung eines krankhaften Versiegens.

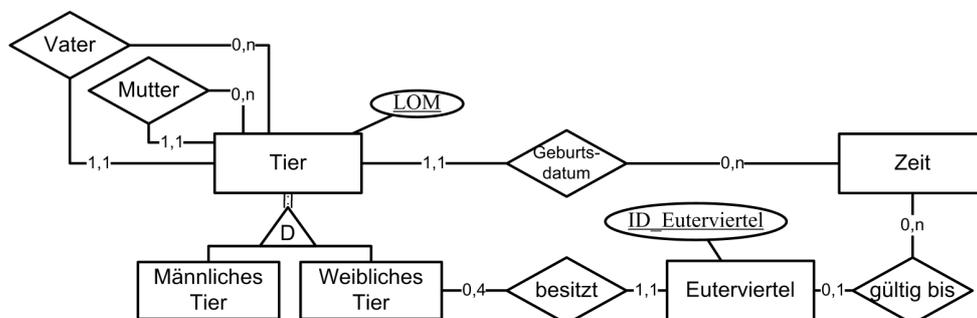


Abbildung 22: E/RM zum Tier

Organisationseinheiten

Neben dem Einzeltier und dem Euterviertel stellen die Tiergruppe, das Stallabteil, der Stall, die Abrechnungseinheit, die Betriebsstätte und das Unternehmen Organisationseinheiten dar (vgl. Kapitel 4.2.2). Ihnen können Leistungen, Kosten und produktionsrelevante Eigenschaften zugewiesen werden. Entsprechend sind sie als eigenständige Entitätstypen zu modellieren. Beziehungen zwischen den Organisationseinheiten existieren in Zugehörigkeiten von Objekten zu anderen Organisationseinheiten, wodurch sich Hierarchien herausbilden (Abbildung 23).

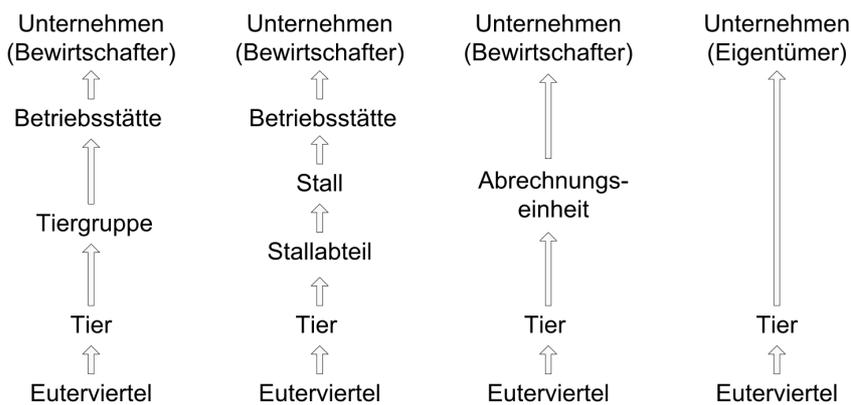


Abbildung 23: Hierarchien zwischen Organisationseinheiten

Die Gültigkeiten der Organisationseinheiten sind zeitlich begrenzt, so dass die Entitätstypen mit einem Zeitstempel versehen werden müssen (Abbildung 24). Eine Ausnahme bildet die Gültigkeit eines Tieres, die als Zugehörigkeit des Tieres zu einer Betriebsstätte über hier nicht dargestellte Entitätstypen für Tierzu- und -abgänge modelliert wird.

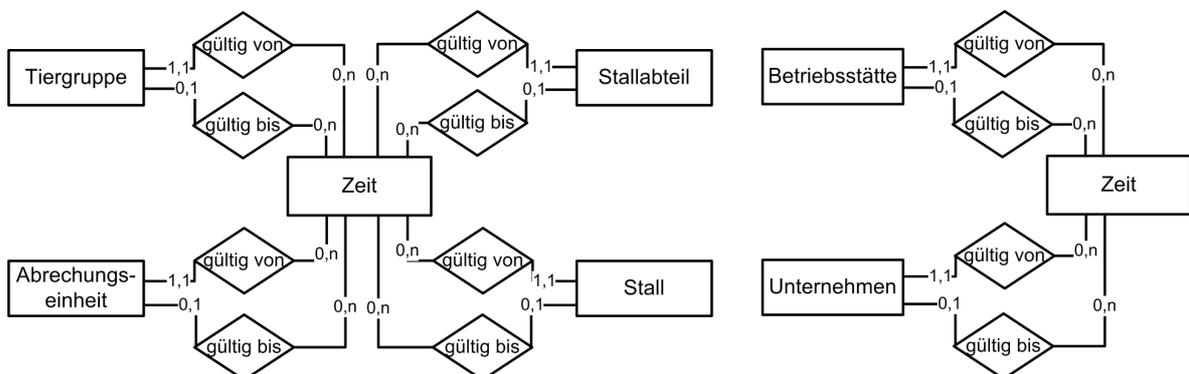


Abbildung 24: E/RM der Gültigkeiten von Organisationseinheiten

Die zeitlich begrenzte Zugehörigkeit einer Organisationseinheit zu einer anderen wird über einen schwachen Entitätstyp mit identifizierenden Beziehungen zum Entitätstyp der niedri-

geren Organisationseinheit und zum Zeit-Entitätstyp zur Zeitstempelung modelliert (Abbildung 25 bis Abbildung 27). Die Zuweisung eines Tieres zu einer Tiergruppe erfolgt beispielsweise über den schwachen Entitätstyp „Tier in Tiergruppe“, der sowohl vom Tier als auch vom Zeit-Entitätstyp identifiziert wird (Abbildung 25). Die Zuordnungen von Stallabteilen zu Ställen, Ställen und Tiergruppen zu Betriebsstätten und Abrechnungseinheiten zu Unternehmen modelliert, da die Zuordnungen selbst zeitunabhängig sind bzw. von der Gültigkeit der Entitäten begrenzt werden. So ist ein Stallabteil beispielsweise baulich ortsfest an einen Stall gekoppelt und kann im Laufe seines Lebenszyklus nicht einem anderen Stall zugeordnet werden.

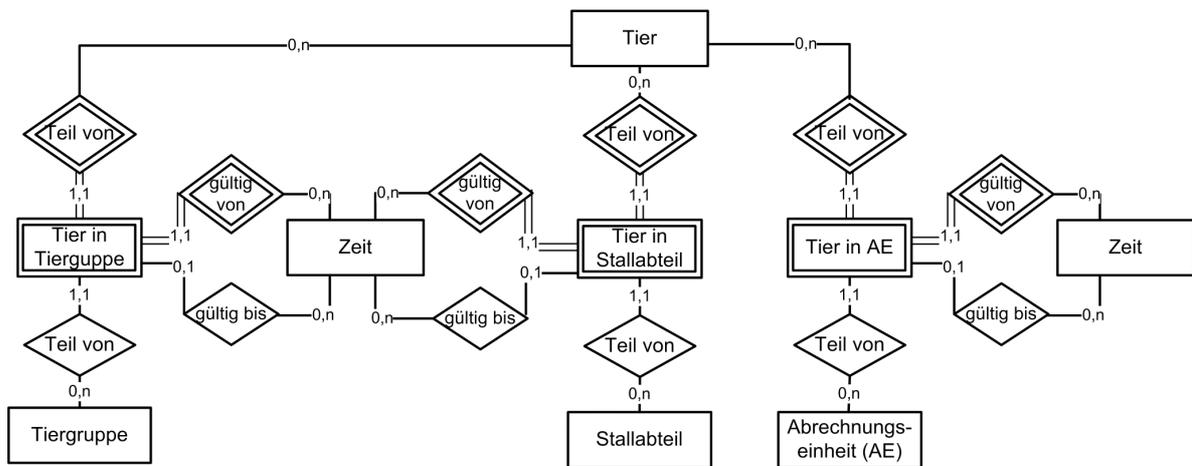


Abbildung 25: E/RM der Zuordnung des Tieres zu den Organisationseinheiten

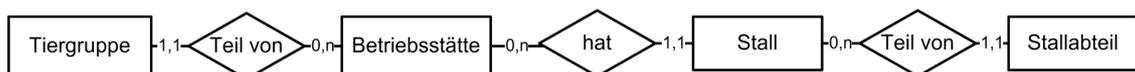


Abbildung 26: E/RM der Zuordnung von Tiergruppen, Betriebsstätten, Ställen und Stallabteilen untereinander

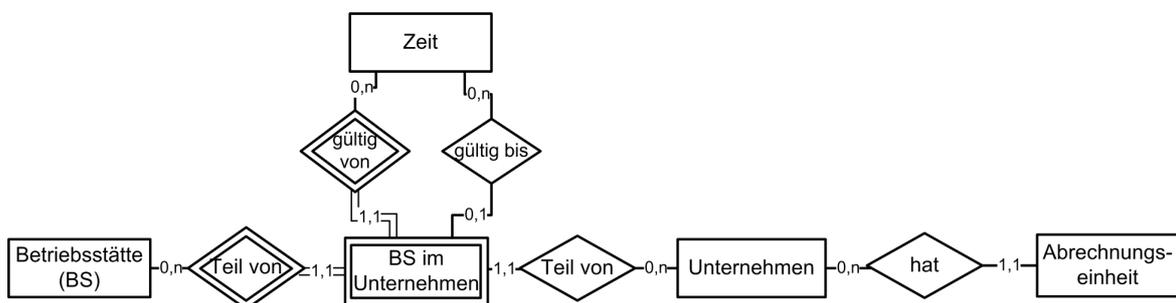


Abbildung 27: E/RM der Zuordnung von Betriebsstätten und Abrechnungseinheiten zu Unternehmen

Die organisatorische Zuordnung von Tieren zu Tiergruppen und die räumliche Zuordnung der Tiere in Stallabteilen stellen einen Zirkelbezug zur Organisationseinheit „Betriebsstätte“ dar (Abbildung 23 bzw. Abbildung 25 und 25). Die Modellierung von Zirkelbe-

zügen birgt die potentielle Gefahr von Inkonsistenzen und sollte deshalb vermieden werden. Unter der Prämisse, dass Tiergruppen nicht zwangsläufig vollständig in einem Stallabteil untergebracht sind, wird keine andere Möglichkeit der Modellierung gesehen: Weder die Beziehungstypen zwischen „Tier“ und „Tiergruppe“, „Tier“ und „Stallabteil“, „Stallabteil“ und „Stall“, „Stall“ und „Betriebsstätte“ und zwischen „Tiergruppe“ und „Betriebsstätte“ sind redundant. Erst die vollständigen Teilpfade ergeben die redundante Zuordnung von Tieren zu Betriebsstätten. Alternativ wäre die Auflösung des Zirkelbezuges durch die Modellierung eines n:m-Beziehungstyps zwischen „Tiergruppe“ und „Stallabteil“ möglich. Der nicht akzeptable Nachteil bei dieser Variante würde jedoch im Verlust der Information bestehen, in welchem Stallabteil sich ein Tier befindet. Im semantischen Schema nicht auflösbare Zirkelbezüge sind durch geeignete Integritätsbedingungen in den nachfolgenden Modellierungsschritten abzusichern.

Krankheitsbehandlungen

Krankheiten stellen als wichtiger direkter und indirekter Kostenfaktor (Behandlungskosten bzw. entgangener Nutzen bei Milch, Fruchtbarkeit etc.) an Bedeutung gewinnende produktionsbeeinflussende Ereignisse in der Milcherzeugung dar. Aus diesem Grund und der Spezifik des Ereignistyps wird die Modellierung von Krankheitsbehandlungen näher erläutert.

Krankheitsbehandlungen stellen ein Ereignis für ein Tier dar und sind als Entitätstyp „Maßnahme“ modelliert. Vergleichbare Maßnahmen finden aber auch an anderen Organisationseinheiten statt. Beispielsweise existiert zwischen der Behandlung einer Milchkuh oder eines Stallabteils (z.B. gegen Parasiten) kein Unterschied in den beschreibenden Merkmalen: Ein Objekt wird zu einer bestimmten Zeit mit einem bestimmten Mittel behandelt. Aus diesem Grund muss in Maßnahmen bei Eutervierteln, Tieren, Tiergruppen, Stallabteilen oder Ställen unterschieden werden. Im E/RM wird der Sachverhalt durch eine Spezialisierung mit entsprechenden Sub-Entitätstypen umgesetzt, die Beziehungstypen mit Entitätstypen für die entsprechenden Organisationseinheiten ausbilden (Abbildung 28). Vererbt werden neben einer eindeutigen ID als Primärschlüssel eine Reihe von Beziehungen zur Charakterisierung einer Maßnahme, wie z.B. der Zeitpunkt, der Typ der Maßnahme, die ausführende Person und eine eventuelle Anwendung von Medikamenten (Abbildung 29). Der Entitätstyp „Medikamentbezug“ spezifiziert dabei die Herkunft des eingesetzten Medikamentes unter anderem als Entsprechung des Medikamentenabgabebesleges.

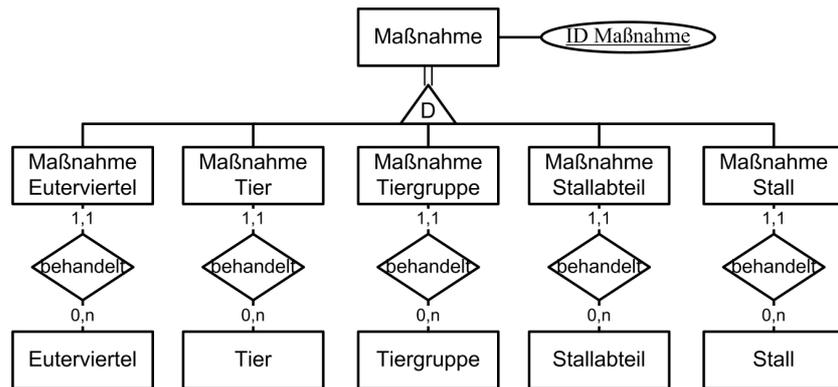


Abbildung 28: E/RM für Spezialisierung von Maßnahmen

Einer Maßnahme geht im Regelfall ein Vorfall oder eine Untersuchung voraus, auch wenn diese teilweise als zusammengehöriges Ereignis betrachtet werden können. Entsprechend ist ein Beziehungstyp zum Entitätstyp „Vorfall/Untersuchung“ modelliert, über den wiederum eine Diagnose abgebildet wird. Auch wenn eine Krankheitsbehandlung stets auf eine Diagnose erfolgt, wird diese jedoch über eine Untersuchung bestimmt. Die Berücksichtigung einer Maßnahme als Folgebehandlung einer vorangegangenen Behandlung wird durch einen selbstreferenzierenden Beziehungstyp „Vorbehandlung“ realisiert. Der Begriff der Selbstreferenz wird hierbei als Beziehung zwischen Entitäten des gleichen Entitätstypen ausgelegt. Nicht der Beziehungstyp referenziert sich selbst, sondern der Entitätstyp referenziert sich über den Beziehungstyp selbst.

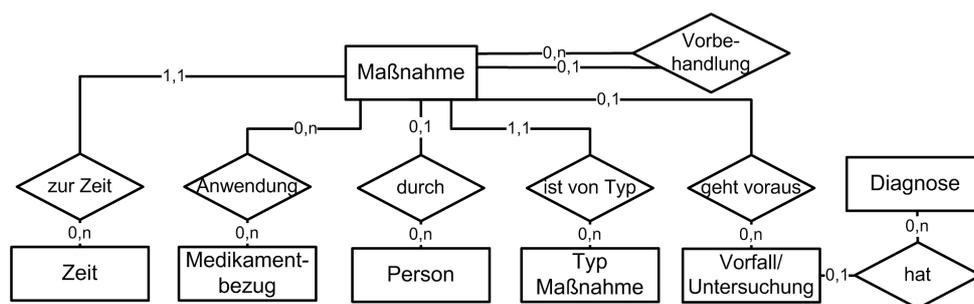


Abbildung 29: E/RM für Maßnahmen

Eine besondere Beachtung gilt der Modellierung von Diagnosen. Die Rinderhaltung ist durch das Fehlen eines standardisierten Diagnoseschlüssels für Krankheiten gekennzeichnet. So existieren unterschiedliche Schlüsselsysteme, die in ihrer Differenziertheit abgebildet werden müssen. Allgemein ist eine Diagnose durch die Angabe des Diagnosesystems, dem die Diagnose angehört, und des Diagnoseschlüssels als eindeutige Identifikation innerhalb des Diagnosesystems beschrieben. In Abbildung 30 ist das resultierende E/RM mit dem Entitätstyp „Diagnose“, dem Primärschlüsselattribut „ID_Diagnose“ als Identifi-

kation über alle Diagnosesysteme hinweg, dem Attribut „Diagnoseschlüssel“ als Identifikation innerhalb eines Diagnosesystems und dem Beziehungstyp zum Typ-Entitätstyp „Diagnosesystem“ dargestellt. Eine hierarchische Organisation der Diagnosen innerhalb eines Diagnosesystems wird über den selbstreferenzierenden Beziehungstyp „übergeordnete_Diagnose“ des Entitätstyps „Diagnose“ abgebildet. Zusätzlich existiert ein Attribut „Diagnoseebene“ für den Entitätstyp „Diagnose“ zur Angabe der Hierarchieebene innerhalb eines Diagnosesystems. Bei paralleler Nutzung verschiedener Diagnosesysteme ist die gegenseitige Zuordnung von Diagnosen verschiedener Diagnosesysteme zueinander notwendig, um Behandlungsdaten über die unterschiedliche Diagnosesysteme auswerten zu können. Diese Diagnosezuordnung wird über einen selbstreferenzierenden Beziehungstyp „Zuordnung_extern“ ermöglicht. Das heißt, der Beziehungstyp „Zuordnung_extern“ bildet Beziehungen zwischen Diagnosen verschiedener Diagnosesysteme und der Beziehungstyp „übergeordnete_Diagnose“ Beziehungen zwischen Diagnosen des gleichen Diagnosesystems ab. Weiterhin sind zur Kennzeichnung anzeige- bzw. meldepflichtiger Krankheiten die Attribute „anzeigepflichtig“ bzw. „meldepflichtig“ vorgesehen.

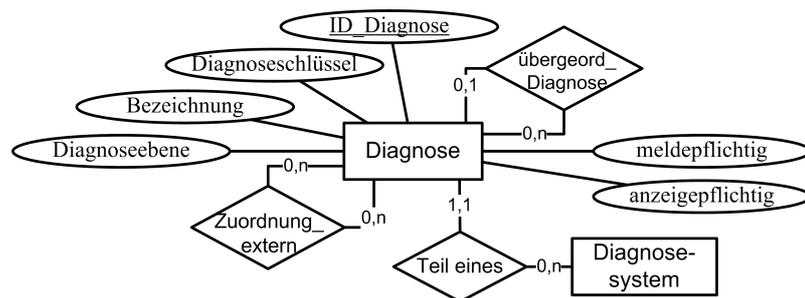


Abbildung 30: E/RM zur Abbildung von Krankheitsdiagnosen

Vererbungshierarchie technischer Objekte

Neben den Organisationseinheiten spielen technische Objekte wie Maschinen oder auch ein Stall oder Stallabteil eine bedeutende Rolle in der Prozessdokumentation. Neben einer zeitlichen Gültigkeit ist ihnen die Beteiligung an gleichen Ereignissen gemein. Dazu zählen beispielsweise die Reinigung, Wartung und Reparatur. Aus diesem Grund ist es sinnvoll, sie unter dem Entitätstyp „Technisches Objekt“ mit einer gemeinsamen Identifikationsnummer („ID_TO“ - ID-Technisches Objekt) zusammenzufassen (Abbildung 31), der in der Modellierung der entsprechenden Ereignisse referenziert werden kann (Abbildung 32). Das gleiche Konstrukt wird auch für die Spezialisierung von Maschinen angewandt (Abbildung 31). Beide Spezialisierungen sind disjunkt und total, das heißt die

Sub-Entitätentypen sind nicht überlappend und alle Entitäten sind einem Sub-Typ zuzuordnen.

Den Nutzen der eingeführten Generalisierungen/Spezialisierungen verdeutlicht Abbildung 32. Der generalisierte Entitätstyp „Technisches Objekt“ kann für die Modellierung beispielsweise einer Reinigung/Desinfektion herangezogen werden, ohne für alle Entitätstypen, die technische Objekte darstellen, gesonderte Reinigungen/Desinfektionen modellieren zu müssen.

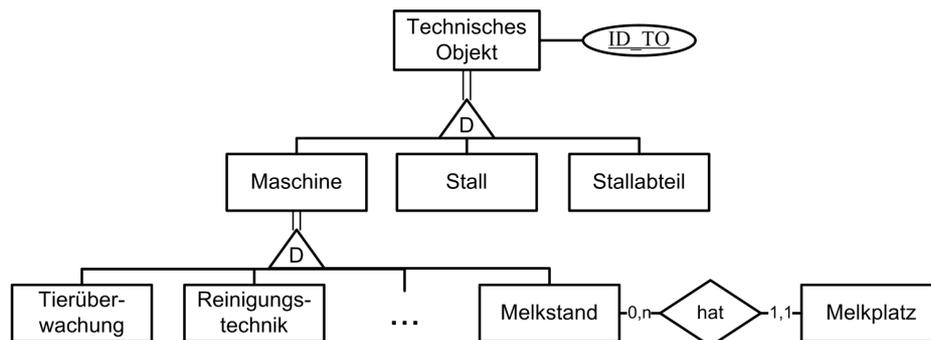


Abbildung 31: E/RM zur Vererbung technischer Objekte

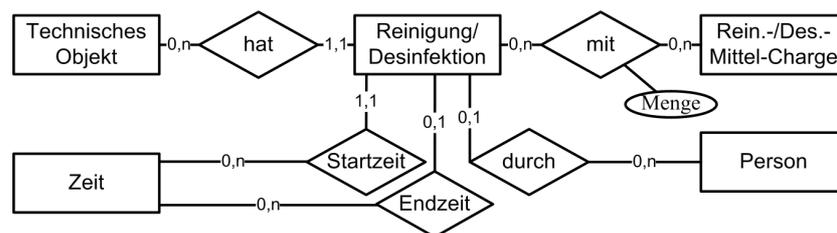


Abbildung 32: E/RM für Reinigung/Desinfektion

Strukturanalogien am Beispiel von Fütterungsdaten

Die Vielzahl an Datenstrukturen abzubildender Prozesse führt zu einem komplexen Datenschema. Bei eingehender Betrachtung zeigen sich jedoch ähnliche Strukturen im Modell, die eine Komplexitätsreduktion ermöglichen. Die dargestellte Modellierung der Krankheitsdaten und technischer Objekte zeigt die Komplexitätsreduktion durch Anwendung des Vererbungs-/Spezialisierungsansatzes auf strukturell identische Sachverhalte. Am Beispiel der Fütterungsdaten zum Einzeltier wird im Folgenden die Zusammenfassung ähnlicher, nicht identischer Modellstrukturen demonstriert.

Die Abbildung von Fütterungsdaten erfordert prinzipiell die Unterscheidung in SOLL- und IST-Daten, das heißt in Daten zur Dokumentation erfolgter Fütterungsvorgänge und in Plandaten. Außer im unterschiedlichen zeitlichen Horizont des Datenanfalls (IST-Daten mehrmals täglich, Plandaten mehrmals im Jahr) unterscheiden sich beide Merkmalskomplexe in weiteren Attributen bzw. in der Semantik der Attribute (Tabelle 4).

Während sich bei den IST-Daten die zeitliche Komponente im Zeitpunkt des Fütterungsvorgangs wieder findet, wird ein SOLL-Datensatz über eine Gültigkeit der Fütterungsanweisung mit Start- und Endzeitpunkt beschrieben. Ebenso reicht die tatsächlich verabreichte Futtermenge des Fütterungsereignisses als Bestandteil der IST-Daten, während bei SOLL-Daten in eine maximale Futtermittelmenge je Abruf (Teilmenge) und eine maximale Tagesmenge unterschieden werden muss. Prinzipiell handelt es sich jedoch um eine ähnliche Merkmalsstruktur: das Tier, der Zeitbezug, das Futtermittel, Mengenangaben und weitere optionale Attribute wie Person und Fütterungstechnik.

Tabelle 4: Merkmalsliste für Fütterungsdaten als Beispiel für Strukturanalogien

IST-Fütterung Einzeltier	SOLL-Fütterung Einzeltier
Tier	Tier
Zeitpunkt der Fütterung	Gültig von Gültig bis
Futtermittel	Futtermittel
Futtermittelmenge	Maximale Futtermittelmenge je Gabe Maximale Futtermittelmenge je Tag
ausführende Person	anweisende Person
Fütterungstechnik	

Während sich bei den IST-Daten die zeitliche Komponente im Zeitpunkt des Fütterungsvorgangs wieder findet, wird ein SOLL-Datensatz über eine Gültigkeit der Fütterungsanweisung mit Start- und Endzeitpunkt beschrieben. Ebenso reicht die tatsächlich verabreichte Futtermenge des Fütterungsereignisses als Bestandteil der IST-Daten, während bei SOLL-Daten in eine maximale Futtermittelmenge je Abruf (Teilmenge) und eine maximale Tagesmenge unterschieden werden muss. Prinzipiell handelt es sich jedoch um eine ähnliche Merkmalsstruktur: das Tier, der Zeitbezug, das Futtermittel, Mengenangaben und weitere optionale Attribute wie Person und Fütterungstechnik.

In Abbildung 33 ist ein E/RM für Fütterungsdaten zum Einzeltier als Kombination von SOLL- und IST-Daten entsprechend den Anforderungen aus Tabelle 4 dargestellt. Die gemeinsame Modellierung erfordert die Einführung eines Unterscheidungsattributes für die Charakterisierung von SOLL- und IST-Daten. Aufgrund der modellweiten Bedeutung eines solchen Unterscheidungsmerkmals wird dazu ein Typ-Entitätstyp „Typ SOLL-IST“ eingeführt. Um eine vollständige Abbildung der Sachverhalte zu gewährleisten, müssen alle Attribute der Merkmalskomplexe im gemeinsamen E/RM enthalten sein. Um jedoch tatsächlich eine Komplexitätsreduktion zu erreichen, kann es sinnvoll sein, logisch gleiche,

aber semantisch abweichende Attribute zusammen zu fassen. Am gewählten Beispiel trifft dies unter anderem auf die Beziehung zum Entitätstyp „Person“ zu, die logisch identisch, semantisch jedoch einerseits die ausführende Person und andererseits die anweisende Person abbildet, je nachdem ob SOLL- oder IST-Daten betrachtet werden (Tabelle 4, Abbildung 33).

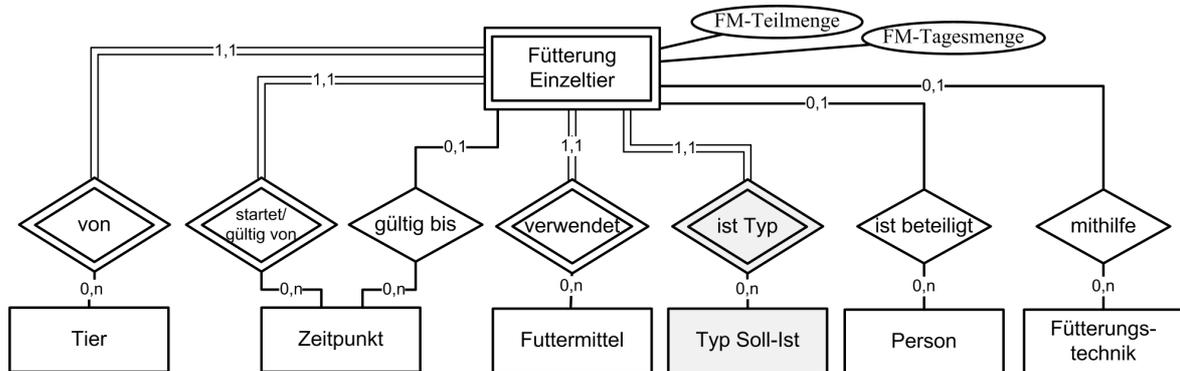


Abbildung 33: E/RM zu Fütterungsdaten des Einzeltieres

Die Beseitigung vergleichbarer Strukturanalogien führt zu einer Generalisierung und damit jedoch zu einer Verringerung der Spezifität von Schemateilen und zu Mehrfachbedeutungen von Schemastrukturen, die dem Konzept eines detaillierten semantischen Entwurfs entgegenstehen. Entsprechend ist im Modellentwurf eine Abwägung zwischen einem semantisch detaillierten und einem komplexitätsreduzierten Entwurf zu treffen.

5.2.2 Semantische Modellierung analytischer Daten

5.2.2.1 Anwendungsspezifische Modellierungsanforderungen

Neben den allgemeinen Kriterien zur Modellierung analytischer Daten (Kapitel 5.1.2) ergeben sich aus der Spezifik des Anwendungsfalles weitere Anforderungen, die an den Modellentwurf gestellt werden.

Wie im vorangegangenen Kapitel 5.2.1.4 dargestellt, existieren Objekttypen, die unter dem Begriff „Organisationseinheiten“ zusammengefasst wurden. Im multidimensionalen Kontext kommt diesen Objekttypen die Rolle von Klassifikationsstufen zu, ihren Beziehungen untereinander die von Klassifikationsbeziehungen. Zu beachten sind dabei jedoch zeitliche Einschränkungen der Ausprägung dieser Klassifikationsbeziehungen und Klassifikationsstufen, die aus ihrer begrenzten Gültigkeit resultieren. Dabei ändert sich hingegen nicht das Schema an sich, sondern nur dessen Ausprägung (vgl. Kapitel 5.1.3.2). Trotzdem wird die Aufnahme einer Kennzeichnung entsprechender zeitlicher Beschränkungen in das Modellschema als wichtig erachtet. Für den Datenbankentwickler hat die Information Auswir-

kungen auf die implementierte Struktur der Datenbank, für den OLAP-Anwendungsentwickler Auswirkungen auf den Entwurf von Dimensionen und Hierarchien und für den Endanwender ist sie für die Interpretation von Daten unerlässlich.

Eine zweite Forderung betrifft die Abbildung von parallelen Dimensionshierarchien. Ausgehend von einer Klassifikationsstufe sind verschiedene Auswertungspfade denkbar und dementsprechend abzubilden. Dies trifft ebenfalls insbesondere auf die Dimension der Organisationseinheiten zu.

Als weitere Anforderung ist die Kennzeichnung von Kennzahleigenschaften im semantischen Modell zu sehen. Auch wenn es sich dabei, wie schon bei den zeitlichen Beschränkungen auf Klassifikationshierarchien, nicht um Strukturinformationen des Schemas handelt, wird ihnen eine hohe Bedeutung für den Schemaentwurf zugeschrieben. Unter Kennzahleigenschaften sind insbesondere die Kennzahldefinition und der Aggregationstyp zu nennen. Die Bedeutung der Kennzahldefinition im semantischen Modell erschließt sich aus der Tatsache, dass im Anwendungskontext für gleich lautende Kennzahlen verschiedene Definitionen existieren und somit schon im Modell deren konkrete Auslegung ersichtlich sein sollte. Aggregationstypen spielen dagegen erst im Entwurf von OLAP-Anwendungen durch Einschränkung der Aggregierbarkeit eine Rolle, gehören jedoch ebenfalls im multidimensionalen Schema vermerkt.

5.2.2.2 Modellierungsnotation

Als Modellierungsansatz wurde das multidimensionale Entity-Relationship-Modell (mE/RM) nach SAPIA et al. (1998) ausgewählt. Vorteile des Ansatzes sind die evolutionäre Erweiterung des E/RM und die intuitive Lesbarkeit der grafischen Repräsentation.

Bekannte Modellelemente

Von SAPIA et al. (1998) wurden als Modellierungselemente eingeführt (Abbildung 34):

- Klassifikationsstufen (dimension level) als spezielle Entitätstypen zur Darstellung von Dimensionsstufen,
- Klassifikationsbeziehungen (rolls-up relationship) als spezielle binäre Beziehungstypen zur Abbildung von Beziehungen zwischen Klassifikationsstufen und
- Faktbeziehungen (fact relationship) als spezielle n -äre¹⁰-Beziehungstypen zwischen Klassifikationsstufen zur Darstellung quantitativer Datenstrukturen mithilfe von
- Kennzahlen als Attribute von Faktbeziehungen.

¹⁰ Der Grad eines Beziehungstypen gibt die Anzahl teilnehmender Entitätstypen an. Ein Beziehungstyp zwischen zwei Entitätstypen (Grad 2) wird binär, zwischen drei Entitätstypen (Grad 3) ternär und zwischen n Entitätstypen folglich n -är genannt.



Abbildung 34: mE/RM-Elemente: Faktbeziehung mit Kennzahl, Klassifikationsstufe, Klassifikationsbeziehung (nach BAUER und GÜNZEL, 2004, S. 167)

Damit sind die in Kapitel 5.1.2 aufgestellten Forderungen zur Modellierung analytischer Daten erfüllt. Parallelhierarchien, wie im vorangegangenen Kapitel beschrieben, können ebenfalls mit diesem Modellierungsansatz abgebildet werden. Die Kennzeichnungen von Kennzahleigenschaften und zeitlichen Einschränkungen auf Klassifikationshierarchien sind jedoch nicht im Ansatz von SAPIA et al. (1998) enthalten.

Zusätzliche Modellelemente

Die fehlende Möglichkeit der Berücksichtigung spezifischer Kennzahl- und Klassifikationspfadsemantik im semantischen analytischen Modell ist als Nachteil zu sehen und wurde durch Erweiterung der existierenden Modellierungselemente von SCHULZE et al. (2005) behoben. Als zusätzliche Elemente wurden ein Eigenschaftsfenster zur direkten Aufnahme spezifischer Kennzahlmerkmale in das Modellschema und eine Kennzeichnung temporaler Restriktionen in der Ausprägung von Klassifikationspfaden eingeführt (Abbildung 35).



Abbildung 35: Erweiterungen der mE/RM-Notation: Eigenschaftsfenster und Kennzeichnung temporaler Restriktion (SCHULZE et al., 2005)

Das Eigenschaftsfenster für Kennzahlen enthält zur eindeutigen Kennzahldefinition beispielsweise Attribute für den Aggregationstyp, um erlaubte Aggregationsoperationen auf der Kennzahl zu definieren (vgl. Kapitel 5.2.2.4), die Metrik, also eine Ableitungsvorschrift der Kennzahl (z.B. als SQL-Code-Fragment), eine erweiterte Beschreibung und die Maßeinheit der Kennzahl.

Das Element zur Kennzeichnung temporaler Restriktionen in der Ausprägung von Klassifikationsstufen und -beziehungen symbolisiert die zeitliche Einschränkung auf Basisgranularität der Zeit, es sei denn, es wird eine Beziehung zu einer abweichenden Granularitätsstufe der Zeit-Dimension gezogen.

5.2.2.3 Ausgewählte Ergebnisse

Analog zu den Modellierungsergebnissen operativer Daten (Kapitel 5.2.1.4) werden im Folgenden die Schemata der multidimensionalen Sichtweise erläutert.

Im Zuge der Darstellung von Dimensionspfaden wird auf die explizite Abbildung der generell höchsten Klassifikationsstufe „ALLE“¹¹ verzichtet. Ihre Existenz wird stillschweigend vorausgesetzt.

Die Dimension „Organisationseinheiten“

Die Bedeutung der Organisationseinheiten als wesentliche Organisationsstruktur des Modells wurde bereits mehrfach verdeutlicht. Im analytischen Entwurf schlägt sich deren Rolle in einer eigenständigen Dimension nieder. Die als Organisationseinheiten bezeichneten Objekttypen werden dabei als Klassifikationsstufen modelliert. Aufgrund der Beziehungen zwischen den Organisationseinheiten, übertragen in Klassifikationsbeziehungen, bilden sich parallele Klassifikationspfade aus, deren Unterscheidungskriterium jeweils verschiedene Gesichtspunkte der Tiergruppierung darstellen (Abbildung 36).

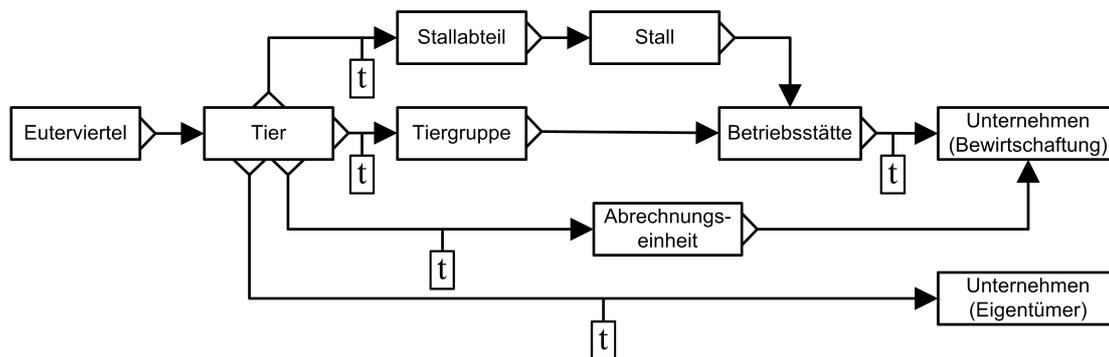


Abbildung 36: mE/RM der Dimension „Organisationseinheiten“ mit parallelen Klassifikationspfaden

Während der Teilpfad Tier→Stallabteil→Stall→Betriebsstätte→Unternehmen→ALLE die räumliche Zuordnung herausstellt, steht beim Teilpfad Tier→Tiergruppe→Betriebsstätte→Unternehmen→ALLE die arbeitsorganisatorische Sicht im Vordergrund. Daneben existieren zwei weitestgehend rechtlich bzw. betriebswirtschaftlich relevante Klassifikationspfade der Organisationseinheiten, der Pfad Tier→Abrechnungseinheit→Unternehmen→ALLE für die Milchablieferung und der Pfad Tier→Unternehmen→ALLE für die Darstellung von Eigentumsverhältnissen.

¹¹ Diese wird auch als „TOP“ oder „ALL“ bezeichnet (SAPIA et al., 1999; GÜNZEL, 2001, S. 48 ff.; LEHNER, 2003, S. 65).

Temporale Restriktionen in der Ausprägung von Klassifikationspfaden wurden mithilfe der entsprechenden Erweiterung der mE/R-Notation modelliert. Auf eine Modellierung durch Anwendung des gleichen Elementes auf Klassifikationsstufen zur Darstellung der zeitlich begrenzten Gültigkeit ihrer Ausprägungen wurde verzichtet. Die begrenzte Gültigkeit eines Objektes in der Klassifikationshierarchie wird stillschweigend vorausgesetzt. Der damit verringerte Informationsgehalt des Modelldiagramms schmälert jedoch nicht den nutzerabhängigen Aussagewert, da es sich nicht um eine entwurfsrelevante Strukturinformation im analyseorientierten Kontext handelt.

Die Dimension „Zeit“

Eine weitere Schlüssel-Dimension resultiert aus der Modellierung der Zeit. Der große Stellenwert der Zeit im analyseorientierten Kontext erhebt die Zeit-Dimension zu einer Standard-Dimension multidimensionaler Schemata, meist in Verbindung mit dem Hierarchie-Pfad Tag→Monat→Quartal→Jahr→ALLE (Abbildung 37). Alternativ existiert ein Parallelpfad Tag→Wochentag→Woche→ALLE, da Wochen nicht eindeutig einem Monat oder Quartal zugeordnet werden können. Die Zuordnung der Kalenderwoche zu einem Jahr ist zwar eindeutig möglich, die aller Tage der letzten bzw. ersten Woche im Jahr jedoch nicht. Aus diesem Grund kann die Klassifikationsstufe „Jahr“ nicht die Rolle der übergeordneten Klassifikationsstufe zur „Woche“ einnehmen.

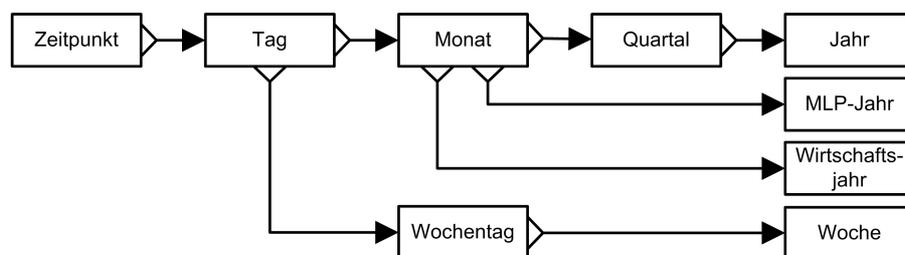


Abbildung 37: mE/RM der Dimension „Zeit“ mit parallelen Klassifikationspfaden

Anwendungsspezifisch sind die Parallelpfade Monat→MLP-Jahr¹² und Monat→Wirtschaftsjahr einzurichten, da das MLP-Jahr abweichend zum Kalenderjahr im Oktober und das landwirtschaftliche Wirtschaftsjahr im Juli beginnt.

Krankheitsbehandlungen

Als Beispiel für die Modellierung eines Datenwürfels unter Einbeziehung der vorgestellten Dimensionen soll die Modellierung von Basis-Kennzahlen zu Krankheitsbehandlungen

¹² MLP - Milchleistungsprüfung

dienen. Es wird direkt Bezug auf die vorgestellte operative Modellierung von Maßnahmen genommen, die die Abbildung von Krankheitsbehandlungen im operativen Modell abdeckt. Eine Faktbeziehung „Behandlungen“ mit den Kennzahlen „Anzahl Behandlungen“ und „Anzahl Erstbehandlungen“ wird eingeführt (Abbildung 38). Beide Kennzahlen besitzen die gleiche Dimensionalität und können deshalb in einer Faktbeziehung modelliert werden. Die Behandlungen, die in die Berechnung der Kennzahl „Anzahl Erstbehandlungen“ eingehen, bilden eine Teilmenge der Behandlungen zur Berechnung der Kennzahl „Anzahl Behandlungen“. Unterscheidungsmerkmal ist der Status einer Behandlung als Folgebehandlung bereits behandelter Erkrankungen. Zu entnehmen ist dies dem Eigenschaftsfenster der Kennzahl, einmal aus der Beschreibung und des Weiteren aus der Metrik in Form eines SQL-Code-Fragmentes. Als weitere Eigenschaften werden beide Kennzahlen über den Aggregationstyp „flow“, also ohne Einschränkungen bezüglich der Aggregierbarkeit definiert (vgl. Kapitel 5.2.2.4).

Relevante Dimensionen sind „Organisationseinheit“, „Zeit“ und „Diagnose“. Auf eine Dimension „Person“ wurde aus Übersichtlichkeitsgründen verzichtet. Ebenso nicht in die Darstellung einbezogen wurden die Klassifikationspfade Monat→MLP-Jahr und Monat→Wirtschaftsjahr der Zeit-Dimension und Tier→Abrechnungseinheit sowie Tier→Unternehmen der Dimension „Organisationseinheiten“. Den vier Pfaden ist gemein, dass sie nicht bewirtschaftungsrelevant in Bezug auf den Gesundheitsstatus von Tieren sind und somit keine Beeinflussung der Kennzahlausprägung zu erwarten ist.

Bisher nicht diskutiert wurde die Diagnose-Dimension. Krankheitsbehandlungen erfolgen grundsätzlich unter Angabe einer Krankheitsdiagnose. Entsprechende Diagnosen sind in Diagnosesysteme eingeordnet, über die feingranulare Diagnosen Krankheitsgruppen zugeordnet sind. Das mE/RM in Abbildung 38 bildet ein 6-Ebenen-Diagnosesystem ab, wie es beispielsweise mit dem Stufenbielschlüssel gegeben ist. Hinzuweisen ist jedoch in zweifacher Hinsicht auf das Problem der Festlegung auf eine Diagnose-Basisgranularität der Faktbeziehung. Zum einen handelt es sich beim Stufenbielschlüssel um eine unbalancierte Klassifikationshierarchie, das heißt nicht alle Krankheitsgruppen besitzen sechs Gliederungsebenen. Zum anderen liegt eine Krankheitsdiagnose aufgrund einer fehlenden medizinischen Notwendigkeit oder Möglichkeit nicht generell in feinsten Granularität vor. Damit enthält das mE/RM für Krankheitsbehandlungen eine Modellierungsansatz bedingte Ungenauigkeit, die Beachtung in den nachfolgenden Entwurfsschritten finden muss.

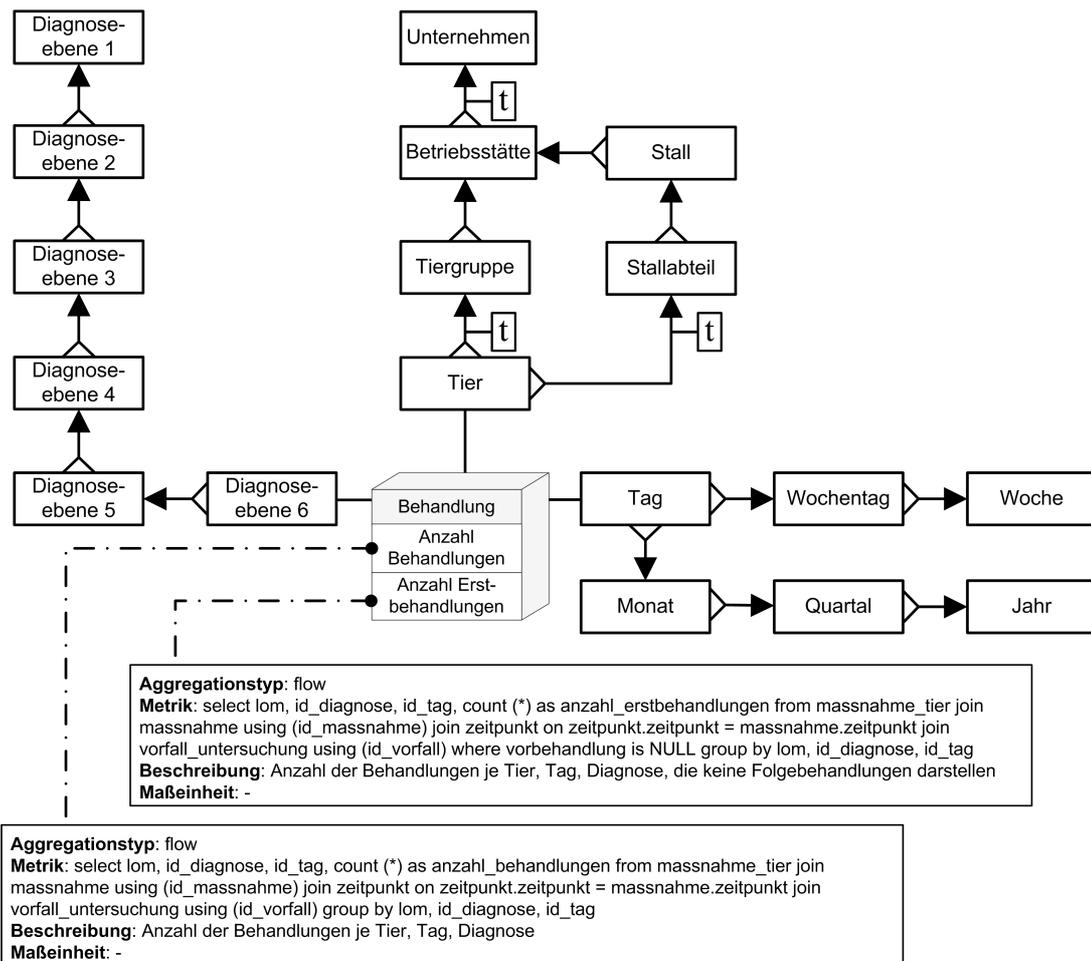


Abbildung 38: mE/RM für Krankheitsbehandlungen

5.2.2.4 Aggregationstypen von Kennzahlen

Eine wichtige Eigenschaft von Kennzahlen stellt deren Summations- bzw. Aggregationstyp¹³ dar. Dieser gestattet Aussagen zu sinnvollen und deshalb erlaubten Aggregationsoperationen auf Kennzahlen bei Nutzung von Navigationsoperationen im multidimensionalen Datenraum (LENZ und SHOSHANI, 1997). Bei LEHNER (2003, S. 67 f.) werden Fakten bzw. Kennzahlen sogar über deren Aggregationstyp definiert. Die drei in der Literatur aufgeführten Aggregationstypen „flow“, „stock“ und „value-per-unit (vpu)“ werden nach einfachen mathematischen Aggregationsoperationen der SQL wie Summe, arithmetisches Mittel, Minimum/Maximum und Anzahl differenziert (LENZ und SHOSHANI, 1997, S. 140 ff.; LEHNER, 2003, S. 71 ff.):

¹³ Beide Bezeichnungen sind hier synonym zu verstehen. Im Folgenden wird jedoch der Begriff des Aggregationstypen verwendet, da der Begriff der Summation irrtümlich mit der Addition als mathematischer Operation assoziiert werden kann.

- **Flow:** Der Aggregationsstyp “flow“ kennzeichnet Merkmale, die frei aggregierbar sind, beispielsweise die Milchmenge.
- **Stock:** Als „stock“ werden Merkmale zusammengefasst, die Zustände wie einen Lagerbestand oder die Anzahl der Tiere eines Stalles beschreiben. Gemeinsames Kennzeichen ist die nicht erlaubte Summenbildung über eine zeitliche Dimension. Die Addition über andere Dimensionen ist erlaubt, beispielsweise die Summe der Tiere aller Ställe einer Betriebsstätte.
- **Value-per-unit (vpu):** Kennzahlen des Aggregationstyps “vpu“ können grundsätzlich nicht addiert werden. Als Beispiele sind der Milchpreis oder die Remontierungsrate zu nennen.

Kennzahlen können jedoch auch nicht-numerische Inhalte wie Texte oder multimediale bzw. nicht quantitative Daten aufweisen (KIMBALL et al., 1998, S. 165; TOTOK und JAWORSKI, 1998, S. 10; BAUER und GÜNZEL, 2004, S. 528). Die Angabe eines nur über mathematische Operationen definierten Aggregationstyps ist für diese Kennzahlen-Klasse aus offensichtlichen Gründen nicht möglich.

Eine sinnvolle und begründete Herangehensweise erscheint in diesem Zusammenhang die Klassifikation nach der Merkmalsskalierung, wie sie aus der Biometrie bekannt ist (KÖHLER et al., 1996, S. 6 ff.; RICHTER, 2004, S. 31 ff.). Die darin vorgenommene Unterscheidung in metrische, ordinal- und nominalskalierte Merkmale bietet sich als grundlegende Klassifikation von Kennzahlen hinsichtlich ihrer Aggregierbarkeit an:

- **Nominal:** Nominalskalierte Merkmale sind in ihrer Ausprägung durch Kategorien gekennzeichnet. Das heißt, es handelt sich um qualitative, nicht metrische Merkmale, wobei die Kategorien gleichwertig sind, also keine Reihenfolge existiert. Als Beispiel sind die Rassen beim Milchrind zu nennen.
- **Ordinal:** Ordinalskalierte Merkmale gehören wie die nominalskalierten zu den qualitativen Merkmalen. Bei ihnen kann jedoch eine Reihenfolge der Kategorien angegeben werden. Beispiele sind Bonitur- oder Schulnoten.
- **Metrisch:** Bei metrischen Merkmalen handelt es sich um quantitative Merkmale, deren Ausprägungen entsprechend auf einer metrischen Skala darstellbar sind. Man unterscheidet in intervall- und proportionalskalierte Merkmale. Auf diese Differenzierung wird jedoch wie bei RICHTER (2004, S. 79) für die Charakterisierung von Kennzahlen im Zusammenhang mit Aggregationsoperationen verzichtet. Typische Beispiele sind die Milchmenge oder die Herdengröße.

Die drei bisher definierten Aggregationstypen „flow“, „stock“ und „vpu“ sind bei Anwendung der biometrischen Systematik eindeutig den metrischen Merkmalen zuzuordnen.

Zur vollständigen Abbildung aller Klassen von Kennzahlen bezüglich ihres Aggregationsverhaltens muss jedoch neben den bisher beschriebenen Aggregationstypen zusätzlich in „nominal“ und „ordinal“ unterschieden werden (Tabelle 5):

- **Nominal:** In Anlehnung an die Merkmalsskalierung der mathematischen Statistik handelt es sich bei dem Aggregationstyp „nominal“ um eine Klasse von Kennzahlen, die qualitative, nicht metrische Merkmale beschreibt.
- **Ordinal:** Der Aggregationstyp „ordinal“ bildet Merkmale ab, die ebenfalls qualitativ, nicht metrischer Art sind, bei denen jedoch eine Rangfolge definiert ist.

Mit dieser Einteilung ist es möglich, die von TOTOK und JAWORSKI (1998, S. 10) angeführten Zeichenketten- und multimedialen Kennzahlen als nominalskaliert und die klassischen numerischen Kennzahlen als metrisch einzuordnen. Die ursprünglich von LENZ und SHOSHANI (1997) vorgenommene Einteilung in „flow“, „stock“ und „vpu“ stellt bei dieser Systematik eine Verfeinerung metrischer Kennzahlen dar (Tabelle 5, Tabelle 6).

Konkrete Anwendung erfährt die Erweiterung der Aggregationstypen bei der Modellierung der Kennzahl des Body-Condition-Score (BCS). Der BCS ist zu den ordinalskalierten Merkmalen zu zählen, das heißt es gibt geordnete Kategorien von 1 bis 5 bzw. 1 bis 9, je nach Bewertungssystem. Einem der drei eingeführten Aggregationstypen „flow“, „stock“ oder „vpu“ ist die Kennzahl nicht zuzuordnen, da bei ihnen der arithmetische Mittelwert erlaubt ist, bei ordinalskalierten Merkmalen wie dem BCS jedoch nicht. Ohne Erweiterung der bestehenden Aggregationstypen ist der BCS als Kennzahl nicht hinreichend genau im konzeptuellen analytischen Modell zu definieren.

Tabelle 5: Aggregierbarkeit von Kennzahlen entsprechend ihres Aggregationstyps (nach LENZ und SHOSHANI [1997] und LEHNER [2003, S. 72], erweitert)

Mathematische Operation	Aggregationstyp				
	metrisch			nominal	ordinal
	flow	stock	value-per-unit		
Anzahl	X	X	X	X	X
Modalwert	X	X	X	X	X
Min/Max	X	X	X	-	X
Zentralwert	X	X	X	-	X
arithmetisches Mittel	X	X	X	-	-
Summe	X	(X)	-	-	-

Tabelle 6: Zuordnung von Aggregationstypen zur statistischen Merkmalskalierung

biometrische Merkmalskalierung	Aggregationstyp				
	metrisch			nominal	ordinal
	flow	stock	value-per-unit		
nominalskaliert	-	-	-	X	-
ordinalskaliert	-	-	-	-	X
metrisch	X	X	X	-	-

5.2.3 Diskussion der semantischen Datenbankentwürfe

Die Unterstützung operativer und analytischer Aufgaben resultiert in zwei getrennten Anwendungssichten auf Daten, die in adäquater Weise ihren Niederschlag im semantischen Datenbankentwurf finden müssen. Ergebnis dieses Entwurfsschrittes sind deshalb zwei, für die operative und die analytische Anwendung getrennte Schemata.

Mithilfe des Entity-Relationship-Ansatzes konnten die anwendungsspezifischen Anforderungen an die Modellierung operativer Daten erfüllt werden. Hervorzuheben ist die Komplexität des resultierenden Datenschemas, für die hauptsächlich drei Gründe identifizierbar sind:

- Der abzubildende Anwendungskontext selbst ist durch einen großen Umfang und eine hohe Komplexität gekennzeichnet.
- Die geforderte Abbildung vergangenheitsbezogener Daten bei Nutzung des Konzeptes der Zeitstempelung von Entitäts- und Beziehungstypen über die Einführung eines schemaweiten Zeit-Entitätstyps führt zu einer erhöhten Anzahl von Entitäts- und Beziehungstypen.
- Die weitestgehend durchgängige Nutzung von Typ-Entitätstypen für mehrfach verwendete Merkmale und vordefinierte Merkmalsausprägungen erhöht ebenfalls die Anzahl der Entitäts- und Beziehungstypen.

Hervorzuheben bleibt weiterhin eine Verschiebung der Semantik von Entitäts- und Beziehungstypen von der Objekt- zu einer Objektversions-Repräsentation aufgrund der Zeitstempelung. Ein Entitätstyp repräsentiert danach nicht eine Menge von Objekten sondern eine Menge von Objektversionen verschiedener Objekte. Gleiches trifft für zeitgestempelte Beziehungstypen zu.

Die Modellierung der Zugehörigkeit von Tieren zu Tiergruppen und der Unterbringung in Stallabteilen führt zu einem parallelen Beziehungspfad zur Betriebsstätte (Tier→Tiergruppe→Betriebsstätte←Stall←Stallabteil←Tier). Dabei handelt es sich jedoch nicht um

redundante Beziehungen, so dass eine Auflösung des Beziehungskreises nicht möglich ist. Entsprechend muss in den nachfolgenden Entwurfsschritten auf eine Integritätssicherung geachtet werden.

Die grundlegende analyseorientierte Sichtweise konnte mit dem multidimensionalen Entity-Relationship-Ansatz in einem semantischen Datenschema abgebildet werden. Als vernachlässigbarer Mangel ist die fehlende direkte Ausweisung von Dimensionen zu benennen, die im Ansatz von SAPIA et al. (1998) nur indirekt aus einer Menge von Klassifikationsstufen und Klassifikationsbeziehungen gebildet werden. Unverzichtbar sind jedoch die bisher fehlende Kennzeichnung temporaler Restriktionen in der Ausprägung von Klassifikationsbeziehungen sowie die Angabe von definierenden und beschreibenden Eigenschaften von Kennzahlen. Um beide Anforderungen erfüllen zu können, wurden eigene Erweiterungen des mE/RM genutzt. Erst die Kennzeichnung besagter temporaler Restriktionen und die Angabe von Kennzahleigenschaften ermöglichen die eindeutige modellseitige Abbildung der Anwendungsfälle. Der Aggregationstyp als beschreibende Kennzahleigenschaft ist bisher jedoch nicht umfassend genug diskutiert. Die in der Literatur beschriebenen Aggregationstypen bzw. nur die Angabe erlaubter Operationen im Schema (wie z.B. im Dimensional Fact-Model [GABRIEL und GLUCHOWSKI, 1998]) werden als nicht ausreichend angesehen. Es wird deshalb vorgeschlagen, sich an der Merkmalsskalierung der Biometrie zu orientieren und die existierenden Aggregationstypen entsprechend einzuordnen, um eine umfassende Aggregationstypen-Definition unter Berücksichtigung nicht proportionalskalierter metrischer Merkmale zu gewährleisten.

6 Logischer Entwurf

Zweiter Schritt des konzeptionellen Datenbankentwurfes ist die logische Datenmodellierung. Neben der Darstellung der grundlegenden Semantik und der Notationselemente liegt der Schwerpunkt dieses Kapitels auf der Transformation semantischer Modellierungselemente, sowohl des E/RM als auch des mE/RM, in logische Strukturen. Ziel des logischen Entwurfes ist die Zusammenführung der getrennten semantischen Entwürfe für operative und analytische Daten in einem gemeinsamen logischen Schema.

Für die logische Datenmodellierung stehen eine Reihe von Modellen zur Verfügung:

- Netzwerkmodell,
- hierarchisches Modell,
- relationales Modell,
- objektorientiertes Modell und
- multidimensionales Modell.

Dem Netzwerk- und dem hierarchischen Modell kommen vor allem eine historische Bedeutung zu, während das relationale Modell (CODD, 1970) das Modell mit der derzeit größten Verbreitung darstellt. Neuere Ansätze basieren auf einer Objektorientierung (objektorientiertes Modell) oder stellen eine direkte Umsetzung der multidimensionalen Sichtweise in einem logischen Datenmodell dar (multidimensionales Modell). Beide Modelle konnten jedoch nicht die Akzeptanz des relationalen Modells erreichen und stellen in diesem Sinne Nischenprodukte dar. Begründet ist dies nicht zuletzt in der weitgehenden Standardisierung des relationalen Modells als auch seiner Datenbanksprache (SQL) und zunehmenden Erweiterungen kommerzieller relationaler Datenbankmanagementsysteme (DBMS) in Richtung einer Objektorientierung bzw. Multidimensionalität. Ergebnis dessen sind z.B. so genannte objektrationale DBMS, die objektorientierte Elemente in ein relationales DBMS integrieren.

Aufgrund der weiten Verbreitung in kommerziellen und freien Datenbankmanagementsystemen und des erfolgreichen Einsatzes sowohl in operativen als auch in analytischen Datenbanksystemen wird das relationale Modell für die nachfolgende Umsetzung der semantischen Entwürfe verwendet.

6.1 Grundlagen des relationalen Modells

Das relationale Modell basiert auf dem Konzept der mathematischen Relation und damit auf der Mengenlehre und Prädikatenlogik. Grundbegriffe des relationalen Modells sind das Attribut, das Tupel, der Primär- und Fremdschlüssel und die Relation, die folgendermaßen definiert sind (ELMASRI und NAVATHE, 2002, S. 225 ff.; SCHUBERT, 2004, S.96 ff., 165 ff.):

- Attribut:** Ein Attribut ist die Bezeichnung einer Menge atomarer Werte mit gleichem Wertebereich.
- Relation:** Eine Relation ist durch einen Relationsnamen und eine Anzahl Attribute gekennzeichnet. Sie enthält eine Menge an Tupeln.
- Tupel:** Ein Tupel ist eine geordnete Liste von Werten der Attribute einer Relation bzw. ein Element des kartesischen Produktes der Attribute einer Relation.
- Primärschlüssel (PK):** Ein Primärschlüssel bezeichnet ein Attribut oder eine Kombination mehrerer Attribute einer Relation, die die Eindeutigkeit der Tupel dieser Relation sicherstellen. Sie dürfen demnach keine gleiche Wertekombination aufweisen bzw. ganz oder in Teilen den NULL-Wert¹⁴ annehmen.
- Fremdschlüssel (FK):** Eine Attributmenge einer Relation r_1 heißt Fremdschlüssel FK auf eine Relation r_2 , wenn jeder Wert von FK einem Wert des Primärschlüssels PK der Relation r_2 entspricht oder den Wert NULL annimmt.

Zur Darstellung logischer Datenbankschemata werden im Folgenden die Notationselemente aus Abbildung 39 für Relationen, Attribute, Fremdschlüsselbeziehungen und für die Kennzeichnung von Primär- und Fremdschlüsseln verwendet. Obligatorische Attribute, also Attribute, die keinen NULL-Wert annehmen dürfen, werden fett gedruckt.



Abbildung 39: Notationselemente zur Darstellung relationaler Schemata

¹⁴ Der Wert NULL drückt einen unbestimmten Zustand eines Attributes aus.

6.2 Abbildung operativer Daten im relationalen Modell

6.2.1 Standardisierte Transformationsregeln für E/RM-Elemente

Die Aufgabe eines logischen Modellierungsschrittes ist die Erstellung eines Datenbankentwurfes in einem logischen Modell. Bei Existenz eines semantischen Entwurfes bedeutet das die Transformation des semantischen Modells in logische Datenstrukturen. Dabei ist es möglich, anhand weitestgehend standardisierter Transformationsregeln für semantische Modellelemente bzw. Elementkombinationen eine Entsprechung im logischen Modell anzugeben (ELMASRI und NAVATHE, 2002, S. 319 ff.). Nachfolgend seien die wichtigsten, in dieser Arbeit verwendeten Transformationsregeln genannt:

- Entitätstypen werden mit ihren Attributen direkt in Relationen und deren Attribute überführt. Als Primärschlüssel der Relation wird ein Schlüssel des Entitätstyps ausgewählt (Abbildung 40).
- Die Transformation von Beziehungstypen richtet sich nach deren Kardinalität.
 - N:m-Beziehungstypen bzw. n-äre Beziehungstypen werden wie in Abbildung 40 dargestellt als eigenständige Relation modelliert. Attribute des Beziehungstyps werden als Attribut der Relation übernommen. Die Primärschlüsselattribute der Relationen der teilnehmenden Entitätstypen sind als Fremdschlüsselattribute in die Relation des Beziehungstyps aufzunehmen. Sie bilden gleichzeitig den Primärschlüssel der Relation.
 - Handelt es sich um einen 1:n-Beziehungstyp, wird der Beziehungstyp nicht in eine eigene Relation überführt, obwohl diese Möglichkeit ebenfalls besteht (ELMASRI und NAVATHE, 2002, S. 321). Ausdruck des Beziehungstyps ist stattdessen das Primärschlüsselattribut der Relation des referenzierten Entitätstyps (n-Seite) als Fremdschlüsselattribut in der Relation des referenzierenden Entitätstyps (1-Seite), wobei sich die Angabe der Kardinalität auf die Maximum-Angabe der min-max-Notation bezieht (Abbildung 41).

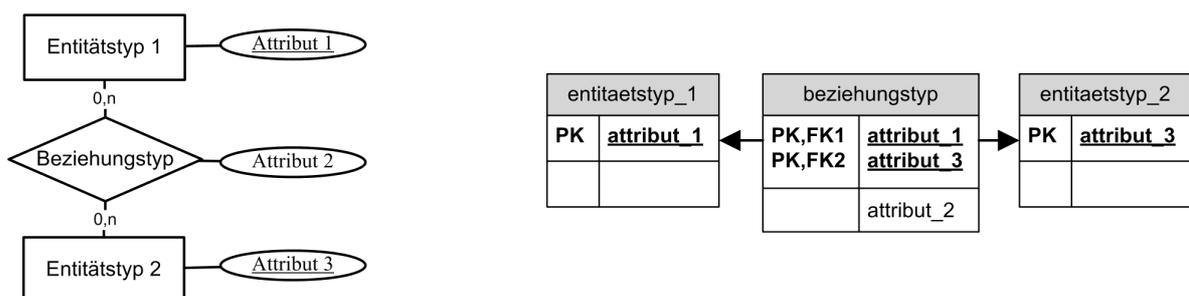


Abbildung 40: Relationale Transformation von Entitätstypen, Attributen und n:m Beziehungstypen

- In ähnlicher Weise wird mit einem schwachen Entitätstyp verfahren. Der schwache Entitätstyp selbst wird in eine eigene Relation umgewandelt. Die Referenz auf den starken Entitätstyp erfolgt dabei als Fremdschlüsselattribut in der Relation des schwachen Entitätstyps. Dieser wird jedoch zusammen mit einem eventuell vorhandenen lokalen Schlüssel zum Primärschlüssel erkoren (Abbildung 41).

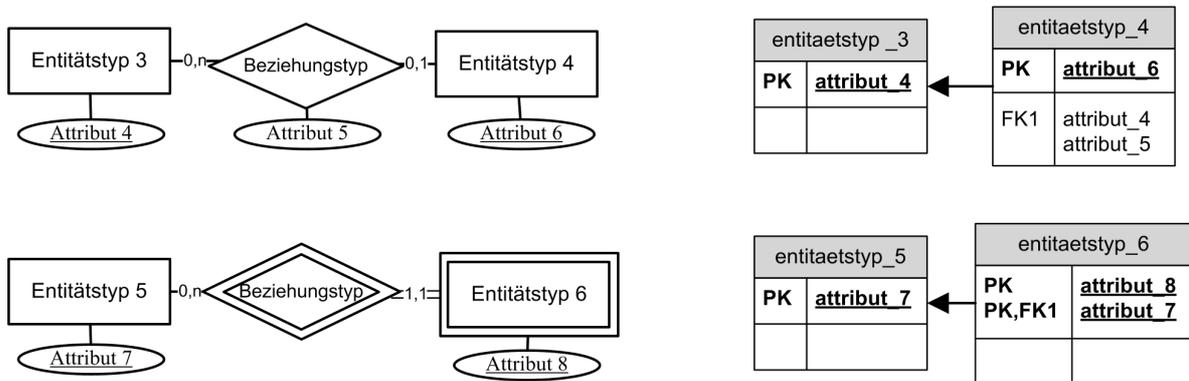


Abbildung 41: Relationale Transformation von 1:n-Beziehungstypen und schwachen Entitätstypen

- Die Umsetzung einer Spezialisierung/Generalisierung ist schwieriger, da mehrere Varianten zur Auswahl stehen. Vorgestellt werden hier drei Varianten, die zur Transformation einer totalen, disjunkten Spezialisierung geeignet sind (Abbildung 42).
 - Im ersten und einfachsten Fall werden alle Entitätstypen, sowohl der Super-Entitätstyp als auch die Sub-Entitätstypen der Spezialisierung in eine einzige Relation überführt (Universalrelation). Primärschlüsselattribut wird das Schlüsselattribut des Super-Entitätstyps. Die Attribute der Sub-Entitätstypen werden als Attribute in die Relation übernommen. Zur Unterscheidung der Tupel der Relation wird ein obligatorisches Typ-Attribut eingeführt, das die Zuordnung der Tupel zu Sub-Entitätstypen abbildet. Diese Variante ist für Sub-Entitätstypen geeignet, die nur wenige spezifische Attribute aufweisen und die Universalrelation deshalb nicht dünn besetzt sein würde.
 - In Variante 2 werden sowohl für den Super-Entitätstyp als auch für die Sub-Entitätstypen eigene Relationen modelliert (vertikale Partitionierung). Die Umsetzung der Sub-Entitätstypen erfolgt dabei adäquat zur Umsetzung schwacher Entitätstypen, so dass die Primärschlüssel der Relationen der Sub-Entitätstypen den Primärschlüssel der Relation des Super-Entitätstyps enthalten. Sinnvoll ist diese Variante, wenn sowohl Anfragen auf den Super-Entitätstyp als auch die Sub-Entitätstypen von Interesse sind, Variante 1 jedoch, z.B. aufgrund vieler spezifischer Attribute der Sub-Entitätstypen, ausscheidet.

3. Im dritten Fall werden nur für die Sub-Entitätstypen eigene Relationen angelegt (horizontale Partitionierung), wobei deren Primärschlüsselattribut dem Schlüssel des Super-Entitätstyps entspricht. Die Anwendung dieser Variante ist angebracht, wenn viele Anfragen auf die Sub-Entitätstypen zu erwarten sind und der Super-Entitätstyp keine Rolle spielt.

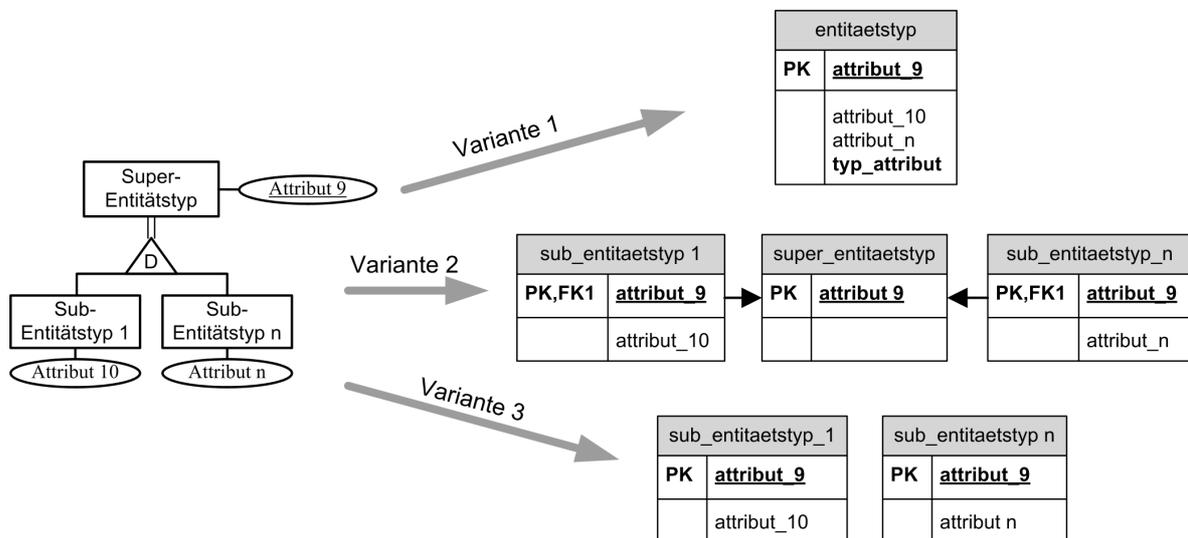


Abbildung 42: Varianten der relationalen Transformation einer Spezialisierung/ Generalisierung

6.2.2 Ausgewählte Ergebnisse der relationalen Modellierung operativer Daten

Entsprechend der vorgestellten E/R-Schemata (Abbildung 22 bis Abbildung 33) für operative Daten werden im Folgenden die relationalen Teilmodelle erläutert, die aus der Anwendung der Transformationsregeln resultieren, wobei nur für das Verständnis des Sachverhaltes notwendige Relationen und Attribute dargestellt werden.

Modellierung des Tieres

Die totale, disjunkte Spezialisierung des Entitätstyps „Tier“ mit den Sub-Entitätstypen für männliche und weibliche Tiere wird in eine einzelne (Universal-)Relation „tier“ überführt (Abbildung 43). Ausdruck der Spezialisierung ist das Attribut „geschlecht“ zur Unterscheidung männlicher und weiblicher Tiere. Die totale Teilnahme an der Spezialisierung drückt sich durch die NICHT-NULL-Bedingung des Attributes aus. Die binären Beziehungstypen „Vater“ und „Mutter“ werden in Fremdschlüsselattribute „lom_vater“ und „lom_mutter“ umgesetzt, die Selbstreferenzen auf die „lom“ der Relation „tier“ darstellen. Das Euterviertel wird als eigenständige Relation mit einer ID als Primärschlüssel modelliert. Das obligate Fremdschlüsselattribut „lom“ als Referenz auf die „lom“ der Relation „tier“ ist Zeichen der Existenzabhängigkeit des Euterviertels vom Tier. Die Sicherstellung, dass nur Tupel für weibliche Tiere Tupel für Euterviertel identifizieren können, muss dabei über

erweiterte Konsistenzsicherungsmechanismen (CHECK-Constraint mit Unterabfrage, Zusicherung, Trigger oder zeitgesteuerte Prüfroutinen) erfolgen (vgl. Kapitel 7.1).

Das Geburtsdatum und die zeitliche Gültigkeit des Euterviertels werden über entsprechende Attribute in den Relationen und Fremdschlüsselbeziehungen zur „zeitpunkt“-Relation abgebildet.

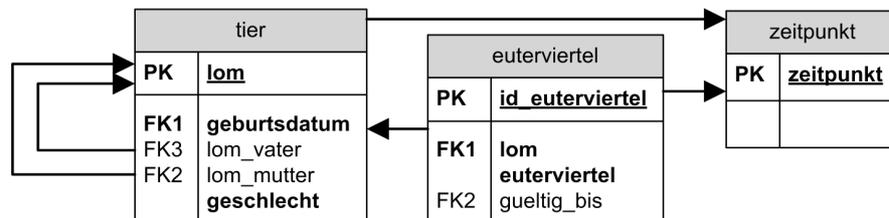


Abbildung 43: Relationales Schema zum Tier

Organisationseinheiten

Die Modellierung der Organisationseinheiten ist eng an die Abbildung von Gültigkeiten und Zugehörigkeiten geknüpft. Generell werden zeitliche Gültigkeiten durch Fremdschlüsselbeziehungen zur Relation „zeitpunkt“ (Entsprechung des Zeit-Entitätstyps) modelliert. Vergleichbar der Handhabung im E/RM werden Gültigkeiten durch Angabe der Attribute „gueltig_von“ und „gueltig_bis“ als Intervallgrenzen dargestellt. Als Beispiel der relationalen Umsetzung der Zeitstempelung von Entitätstypen zeigt Abbildung 44 die Modellierung von Gültigkeiten der Organisationseinheiten. Wie beschrieben werden dazu in jeder Relation Attribute für die Intervallgrenzen angelegt, die über Fremdschlüsselbeziehungen auf die Relation „zeitpunkt“ verweisen. Zusammen mit dem Attribut für die Bezeichnung der Organisationseinheit bildet die untere Intervallgrenze einen mehrteiligen Schlüsselkandidaten aus, der aus diesem Grund auch mit einer NICHT-NULL- und einer UNIQUE-Bedingung belegt werden muss. Als Primärschlüsselattribut wird jedoch eine ID zur Vermeidung von „Schlüsselkaskaden“ (Kapitel 5.2.1.3) eingeführt.

Zur Absicherung der disjunkten zeitlichen Gültigkeit von Organisationseinheiten, das heißt der Bedingung, dass Versionen einer Organisationseinheit nicht zeitlich überlappen, sind Konsistenzsicherungsmechanismen zu nutzen (Kapitel 7.1). Gleiches gilt für die Bedingung, dass der Attributwert des Intervall-Anfanges in einem Tupel stets kleiner als der Attributwert des Intervall-Endes ist.

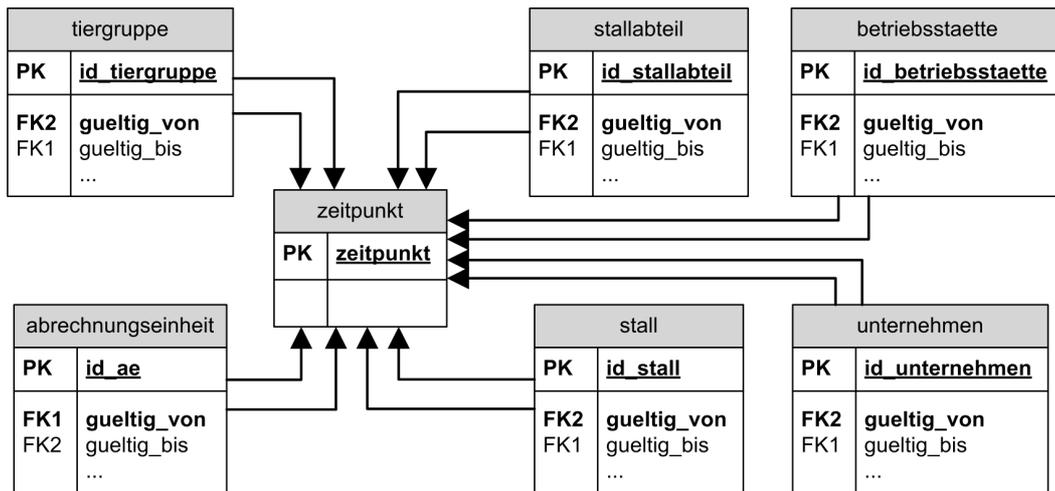


Abbildung 44: Relationales Schema der zeitlichen Gültigkeiten von Organisationseinheiten

Die Verfahrensweise bei der relationalen Modellierung der Zugehörigkeiten von Organisationseinheiten ist ähnlich jener von zeitgestempelten Entitätstypen. Die Zugehörigkeiten, die semantische Beziehungen darstellen, wurden aufgrund der Schlüsselproblematik bei einer Zeitstempelung (Kapitel 5.1.3.1) als schwache Entitätstypen modelliert. Entsprechend werden Relationen angelegt, deren komposite Primärschlüssel aus Fremdschlüsseln auf die Primärschlüssel der Relationen der starken Entitätstypen bestehen (Abbildung 45 bis Abbildung 47). Beispielsweise wird die Zugehörigkeit von Tieren zu Tiergruppen über eine Relation „tier_in_tiergruppe“ abgebildet, deren Primärschlüssel sich aus der „lom“ des Tieres und der unteren Intervallgrenze „gueltig_von“ zusammensetzt. Der Unterschied zur vorangegangenen Modellierung der Gültigkeitszeiten besteht demnach nur in der Nicht-Verwendung einer ID als Primärschlüssel. Die Aussagen zur Konsistenzsicherung gelten jedoch in gleicher Weise. Die nicht zeitgestempelten Zugehörigkeiten von Stallabteilen zu Ställen, Ställen und Tiergruppen zu Betriebsstätten und Abrechnungseinheiten zu Unternehmen, im E/RM als 1:n-Beziehungstypen modelliert, werden entsprechend den Transformationsregeln mit einem Fremdschlüsselattribut in der Relation des referenzierenden Entitätstyps umgesetzt (Abbildung 46 und Abbildung 47). Die funktionale Abhängigkeit wird durch die NICHT-NULL-Bedingung des Fremdschlüsselattributes ausgedrückt.

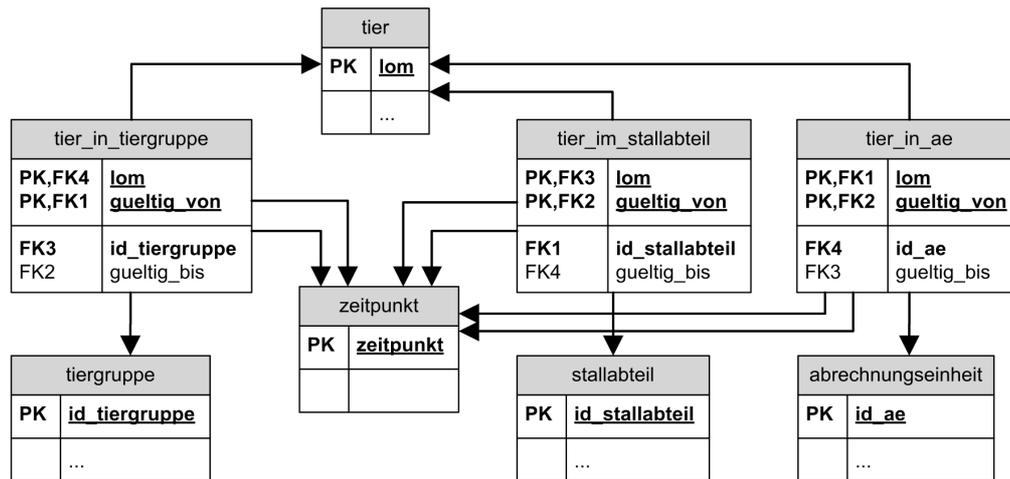


Abbildung 45: Relationales Schema für Zugehörigkeiten des Tieres

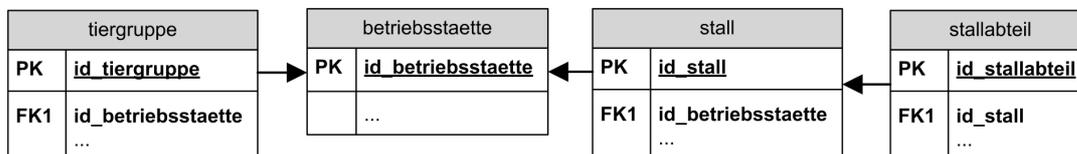


Abbildung 46: Relationales Schema für Zugehörigkeiten von Tiergruppen, Stallabteilen und Ställen zu Betriebsstätten

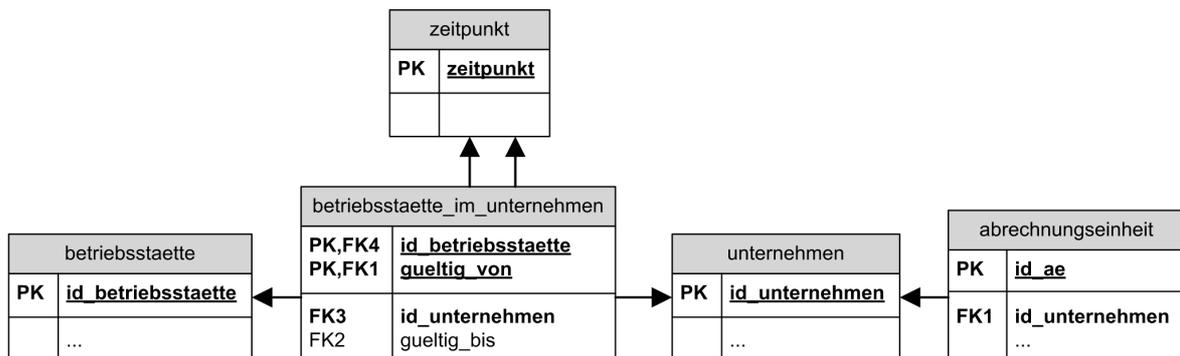


Abbildung 47: Relationales Schema für Zugehörigkeiten von Betriebsstätten und Abrechnungseinheiten zu Unternehmen

Krankheitsbehandlungen

Die totale, disjunkte Spezialisierung einer Maßnahme wird, anders als beim Tier vorgestellt, über eine vertikale Partitionierung relational umgesetzt, also mit einer Relation für den Super-Entitätstyp als auch mit Relationen für jeden Sub-Entitätstypen (Abbildung 48). Vererbter Primärschlüssel für alle Subtypen ist die ID für eine Maßnahme der Relation des Super-Entitätstyps. Die 1:n-Beziehungstypen der Sub-Entitätstypen mit den jeweiligen Entitätstypen der Organisationseinheiten werden wie gewohnt in einen Fremdschlüssel (NICHT-NULL-Bedingung) in die spezialisierten Maßnahme-Relationen transformiert. Im Rahmen der Spezialisierung werden jedoch keine Nicht-Schlüsselattribute und Bezie-

lungen des Super-Entitätstyps „Maßnahme“ an die Relationen der Sub-Entitätstypen vererbt. Die Merkmale einer Maßnahme werden somit als Attribute der Relation „massnahme“ modelliert (Abbildung 49). Hinzuweisen ist dabei auf die NICHT-NULL-Bedingung der Attribute „zeitpunkt_massnahme“ und „bezeichnung_massnahme“ als notwendige spezifizierende Merkmale einer Maßnahme. Ebenfalls hervorzuheben ist das Fremdschlüsselattribut „vorbehandlung“, das über eine Fremdschlüsselbeziehung auf die „id_massnahme“ der gleichen Relation verweist. Der im E/RM auftretende n:m-Beziehungstyp für eine mögliche Medikamentanwendung wird adäquat der Transformationsregeln in eine eigene Relation mit einem Primärschlüssel, bestehend aus den Primärschlüsselattributen der Relationen der teilnehmenden Entitätstypen modelliert.

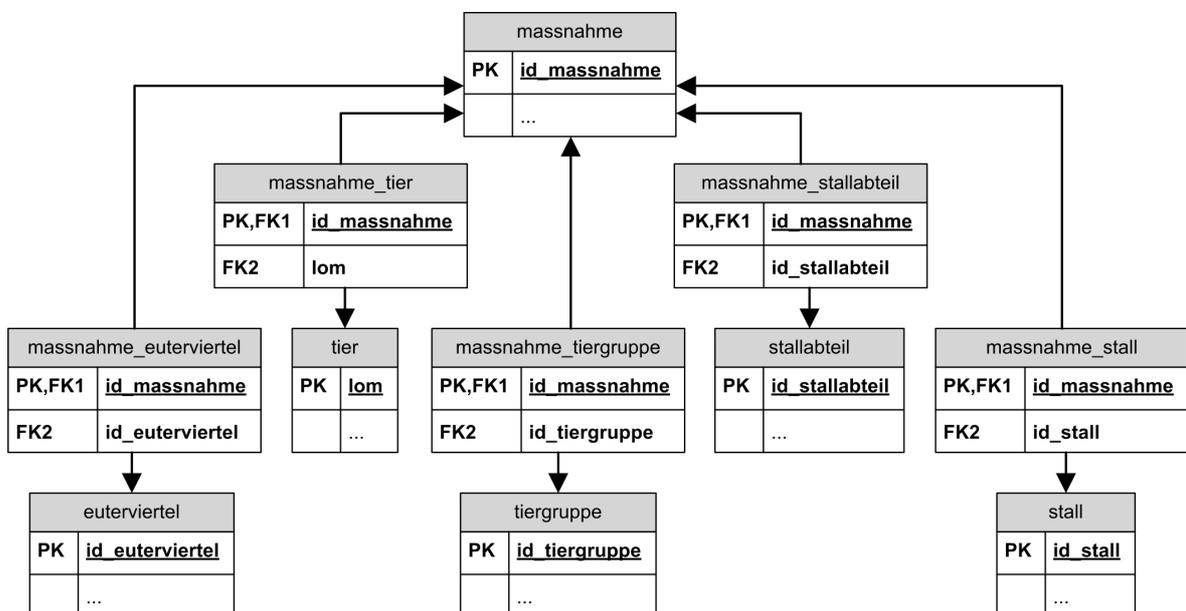


Abbildung 48: Relationales Schema für die Spezialisierung einer Maßnahme

Ausgehend von der Spezialisierung einer Maßnahme für verschiedene Organisationseinheiten im E/RM existieren im relationalen Schema zusammenfassend eine Relation „massnahme“, die die Merkmale der Maßnahme abbildet, und mehrere, diese Relation referenzierende Relationen zur Spezifizierung der Organisationseinheit, an der die Maßnahme ausgeführt wird. Im Zusammenhang der Konsistenzsicherung ist darauf zu achten, dass jede „id_massnahme“ nur in einer Maßnahme-Relation der Organisationseinheiten referenziert wird.

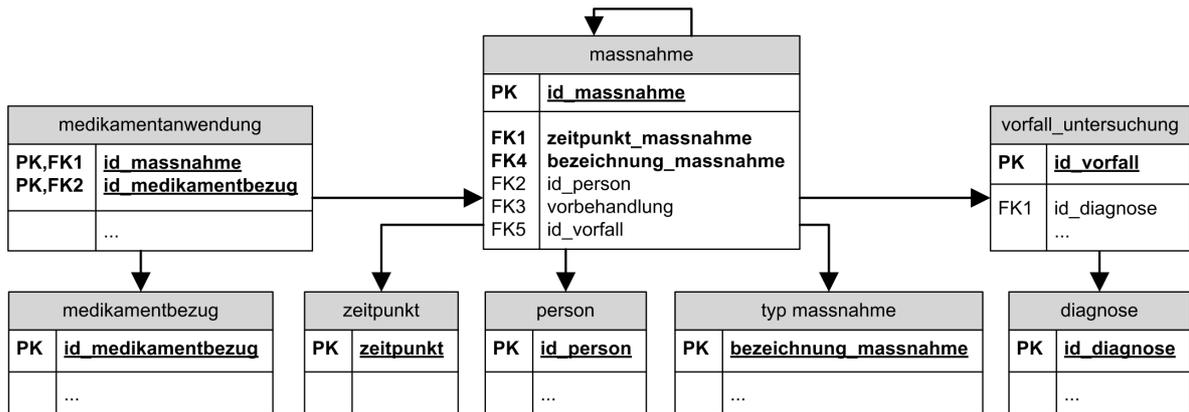


Abbildung 49: Relationales Schema für Maßnahmen

Die relationale Umsetzung des E/RM für Diagnosen ist in Abbildung 50 zu sehen. Primärschlüsselattribut der Relation „diagnose“ ist auch hier eine ID. Für die Merkmale Diagnoseschlüssel, Bezeichnung der Diagnose und das Diagnosesystem sind gleichnamige, obligatorische Attribute modelliert, wobei das Attribut „diagnosesystem“ durch eine Fremdschlüsselbeziehung auf die Relation „diagnosesystem“ gekennzeichnet ist. Das Attribut „uebergeordnete_diagnose“ ist ebenfalls Fremdschlüsselattribut, jedoch auf die ID der eigenen Relation. Dadurch ist es möglich, Hierarchien innerhalb von Diagnosesystemen abzubilden. Die Zuordnung von Diagnosen verschiedener Diagnosesysteme erfolgt in einer eigenen Relation „zuordnung_extern“ mit Fremdschlüsselattributen auf die „id_diagnose“ der Relation „diagnose“, wodurch eine Diagnose in mehrere Diagnosesysteme umgeschlüsselt werden kann. Grundsätzlich ist wiederum über geeignete Mechanismen sicherzustellen, dass Werte im Attribut „uebergeordnete_diagnose“ tatsächlich Diagnosen des gleichen Diagnosesystems und Werte des Attributes „id_umgeschlüsselt“ der Relation „zuordnung_extern“ Diagnosen eines verschiedenen Diagnosesystems referenzieren. Ein Beispiel für die Ausprägung der Schemata zu Krankheitsdiagnosen stellt Abbildung 51 dar.

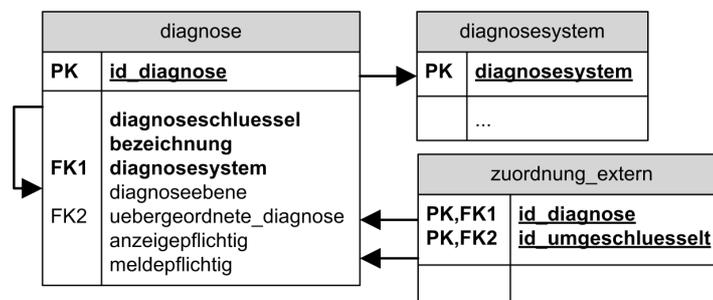


Abbildung 50: Relationales Schema für Diagnosen

diagnose	<u>id_diagnose</u>	<i>diagnose schluessel</i>	<i>bezeichnung</i>	<i>diagnose system</i>	<i>uebergeordnete_diagnose</i>	...
	1	1.	Organkrankheiten	Staufenbiel	NULL	
	2	1.13.	Mastitis	Staufenbiel	1	
	9	2	Eutererkrankungen	Herde	NULL	
	11	24	Mastitis	Herde	9	

diagnosesystem	<u>diagnosesystem</u>
	Staufenbiel
	Herde

zuordnung_extern	<u>id_diagnose</u>	<u>id_umgeschluesselt</u>
	9	1
	11	2

Abbildung 51: Beispiel für die Ausprägung der Relationsschemata „diagnose“, „diagnosesystem“ und „zuordnung_extern“

Vererbungshierarchie technischer Objekte

Wie im vorangegangenen Beispiel der Maßnahmen werden die Spezialisierungen von technischen Objekten und Maschinen über je eine Relation für den Super-Entitätstyp und je eine Relation für die Spezialisierung dargestellt (Abbildung 52). Vererbtes Primärschlüsselattribut ist eine ID. Die Relation „maschine“, die gleichzeitig Relation eines Subals auch eines Superentitätstyps ist, vererbt die ID an die Relationen ihrer Sub-Entitätstypen weiter.

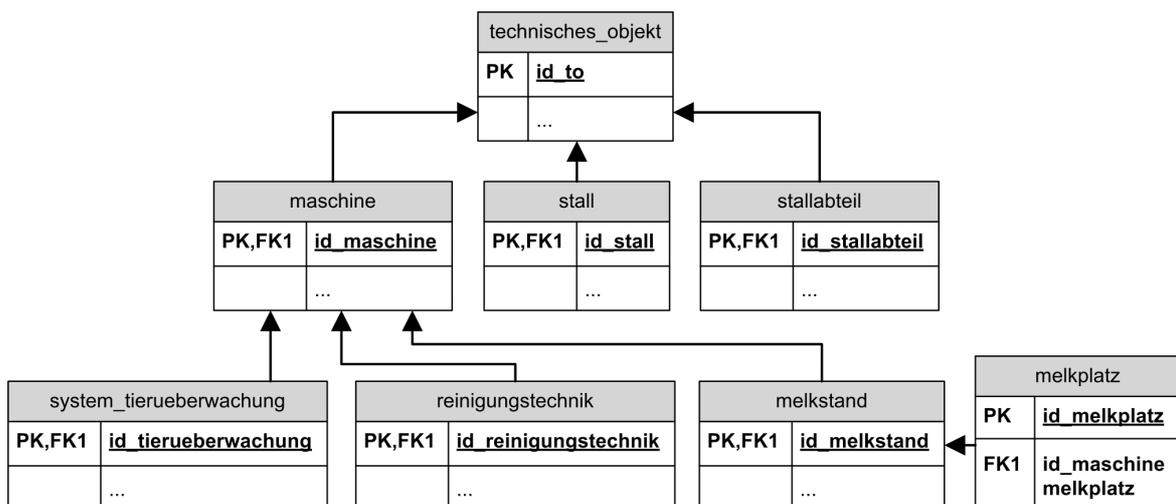


Abbildung 52: Relationales Schema für die Spezialisierung von technischen Objekten

Das ermöglicht es, Ereignisse an technischen Objekten (inklusive Maschinen) über eine einzelne ID zu modellieren. Am Beispiel einer Reinigung/Desinfektion wird die Möglichkeit in Abbildung 53 demonstriert. Die Reinigung/Desinfektion als eigenständige Relation referenziert die ID eines technischen Objektes, so dass die Relation gleichermaßen Reinigungen/Desinfektionen von Ställen als auch von Melkständen abbilden kann.

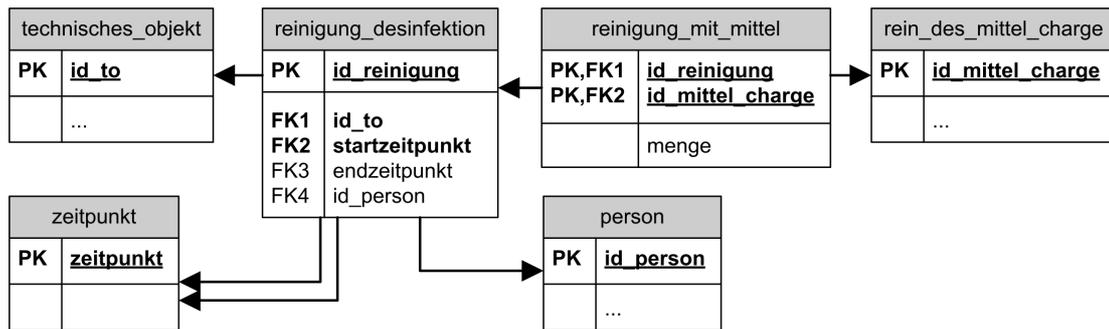


Abbildung 53: Relationales Schema für Reinigungen/Desinfektionen

Strukturanalogien am Beispiel von Fütterungsdaten

Am Beispiel der Fütterungsdaten des Einzeltieres wurde die Komplexitätsreduktion im E/RM durch Zusammenfassung ähnlicher Schemastrukturen erläutert. In Abbildung 54 ist dazu das entsprechende relationale Schema dargestellt. Der schwache Entitätstyp „fütterung_einzeltier“ bildet eine gleichnamige Relation aus, deren Primärschlüssel sich aus den Primärschlüsselattributen der Relationen der identifizierenden Entitätstypen zusammensetzt. Die Folge ist ein 4-teiliger Primärschlüssel. Neben der LOM des Tieres und dem Zeitpunkt wird ein Fütterungsereignis eindeutig durch die Angabe des Futtermittels charakterisiert. Durch die Kombination von SOLL- und IST-Daten zur Fütterung in einer Relation wird jedoch zusätzlich der Typ eines Tupels zur Sicherung der Eindeutigkeit benötigt („soll_ist“).

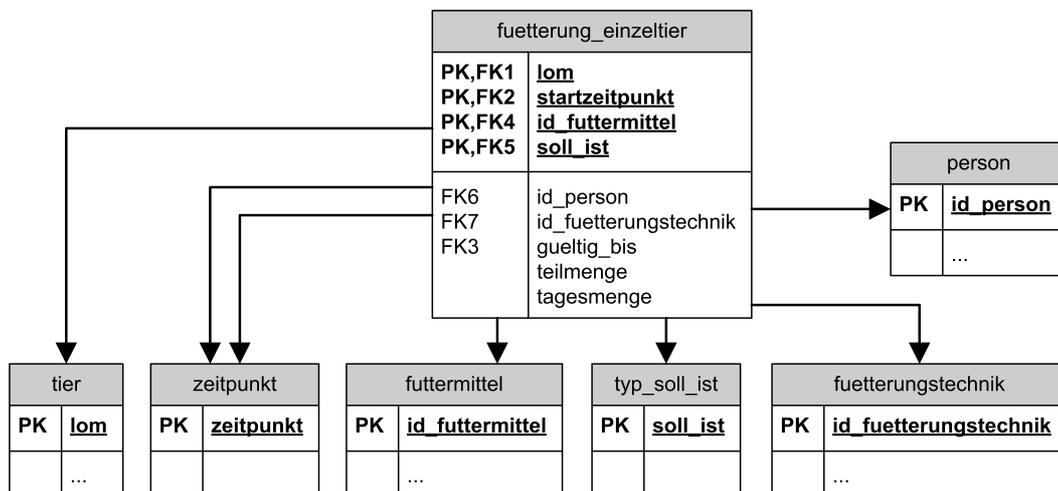


Abbildung 54: Relationales Schema für die Fütterung von Einzeltieren

Ebenfalls durch die gemeinsame Modellierung sind die Werte einiger Attribute unterschiedlich zu interpretieren. So gibt der Wert des Attributes „startzeitpunkt“ bei einem IST-Daten-Tupel den Zeitpunkt der Fütterung an, während die Werte desselben Attributes bei SOLL-Daten als untere Intervallgrenze der Gültigkeit zu betrachten sind. Ähnliche

Interpretationsunterschiede treten, wie schon bei der Erläuterung des E/RM genannt, bei den Attributen „id_person“, „teilmenge“ und „gesamtmenge“ auf.

6.3 Abbildung analytischer Daten im relationalen Modell

6.3.1 Standardisierte Transformationsregeln für mE/RM-Elemente

Wie für die Elemente des Entity-Relationship Modells lassen sich auch für das multidimensionale E/RM Transformationsregeln angeben, die eine Überführung entsprechender Modellelemente in relationale Strukturen ermöglichen (BAUER und GÜNZEL, 2004, S. 203 ff.). Klassifikationsstufen werden wie Entitätstypen des E/RM als eigenständige Relationen modelliert (Abbildung 55). Klassifikationsbeziehungen zwischen ihnen wurden als spezieller 1:n-Beziehungstyp des E/RM eingeführt und werden deshalb auch wie dieser mithilfe eines Fremdschlüssels umgesetzt. Das heißt, die Relation der Klassifikationsstufe niedriger Granularität erhält ein Fremdschlüsselattribut als Referenz auf das Primärschlüsselattribut der Relation der Klassifikationsstufe nächst höherer Granularität, um die funktionale Abhängigkeit abzubilden.

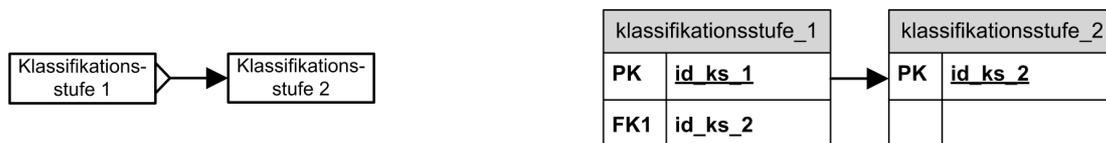


Abbildung 55: Relationale Transformation von Klassifikationsstufen (KS) und Klassifikationsbeziehungen

Zur Überführung der Faktbeziehung in relationale Strukturen wird ebenfalls von deren Definition als Erweiterung des E/RM ausgegangen. Eingeführt wurde die Faktbeziehung als spezieller n-ärer Beziehungstyp und wird deshalb adäquat im relationalen Modell behandelt. Für jede Faktbeziehung wird eine eigene Relation angelegt (Abbildung 56). Kennzahlen stellen Attribute der Faktbeziehung dar und werden gleichermaßen als Attribut der Relation modelliert. Die Klassifikationsstufen, die eine Faktbeziehung ausbilden, gehen über Fremdschlüsselbeziehungen in entsprechenden Fremdschlüsselattributen in die Relation ein. Diese bilden dabei zusammen gleichzeitig den Primärschlüssel aus.

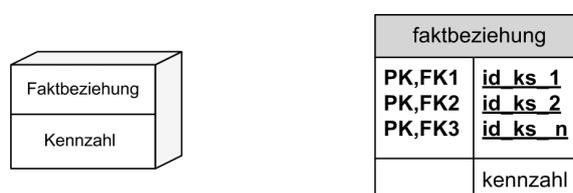


Abbildung 56: Relationale Transformation einer Faktbeziehung

6.3.2 Transformationsregeln für die mE/RM-Erweiterungen

Auf gleichem Wege wie die bekannten mE/RM-Elemente sind auch die Erweiterungen des mE/RM ins relationale Modell zu überführen.

Die Kennzeichnung temporaler Restriktionen in der Ausprägung von Klassifikationsstufen und -beziehungen ist als zeitliche Gültigkeit entsprechender Elementinstanzen definiert worden. Auf das E/RM Bezug nehmend, werden zeitliche Gültigkeiten als Beziehung zu einem Zeit-Entitätstyp bzw. auf relationaler Ebene zu einer Zeit-Relation modelliert. Dies erfolgt auf Ebene der niedrigsten Zeit-Granularität, wenn keine anderweitige Angabe vorliegt. Für die Transformation temporaler Restriktionen von Klassifikationsbeziehungen bieten sich zwei verschiedene Möglichkeiten an.

Die erste Variante entspricht grundsätzlich der ursprünglichen Umsetzung einer Klassifikationsbeziehung durch einen Fremdschlüssel als Referenz auf die Relation der nächst höheren Klassifikationsstufe (Abbildung 57).

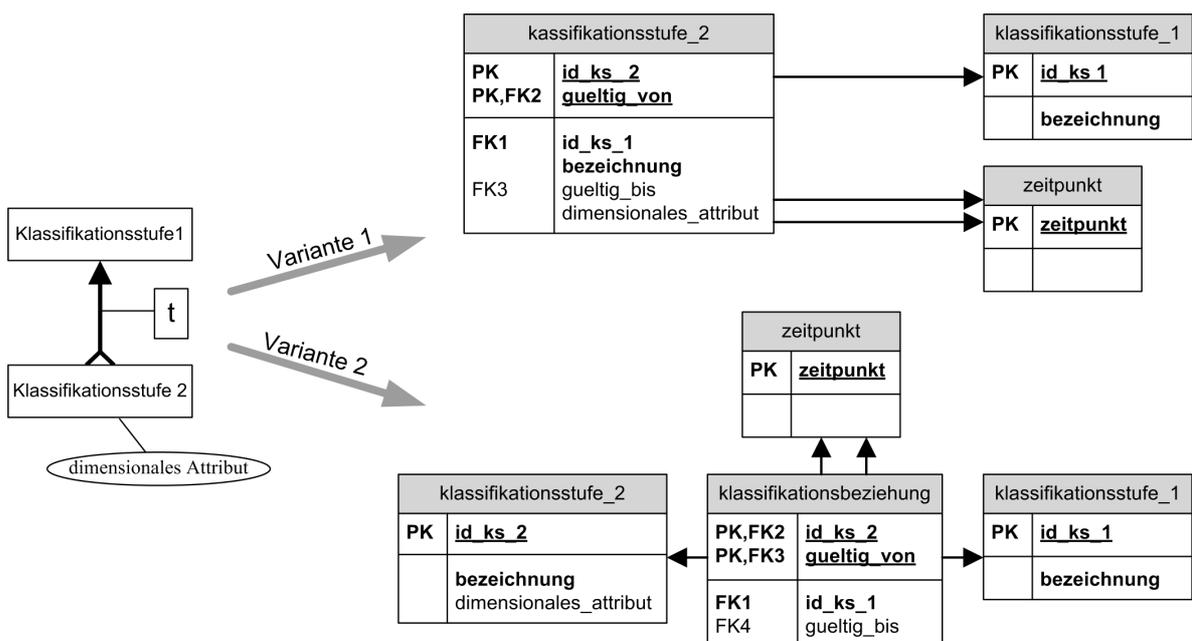


Abbildung 57: Relationale Transformation der temporalen Restriktion einer Klassifikationsbeziehung

Als Ausdruck der zeitlich begrenzten Gültigkeit werden jedoch die Zeit-Attribute zur Abbildung der Gültigkeit als Fremdschlüsselattribute in die Relation der niedrigen Klassifikationsstufe und die identifizierende Zeit in den Primärschlüssel aufgenommen. Damit wird die zeitliche Gültigkeit der Klassifikationsbeziehung auf die niedrige Klassifikationsstufe übertragen. Die funktionale Abhängigkeit der Klassifikationsstufen bleibt dadurch in ihrer Bedeutung und vor allem in der relationalen Struktur erhalten. Nachteilig wirkt sich die Art der Modellierung bei vorhandenen dimensional Attributen der niedrigen Klassi-

fikationsstufe aus. Bei Änderung der Klassifikationsbeziehung werden die Werte der dimensional Attribute redundant gespeichert. Des Weiteren muss die Auslegung der Klassifikationsstufen-ID überdacht werden. Wird sie, wie in Abbildung 57 dargestellt, als ID eines Klassifikationsknotens betrachtet, drohen Primärschlüsselkaskaden wie unter Kapitel 5.2.1.3 beschrieben, wenn der Klassifikationspfad durch mehrere zeitabhängige Klassifikationsstufen oder -beziehungen gekennzeichnet ist. Besser ist die Auslegung der ID als Identifikation der Version eines Klassifikationsknoten inklusive seiner Klassifikationsbeziehung. Dadurch entfällt die zeitliche Gültigkeit im Primärschlüssel.

Die zweite, allgemeinere Lösung besteht in der Einführung einer eigenständigen Relation für die Klassifikationsbeziehung (Abbildung 57). Die zeitliche Gültigkeit wird, wie bei der operativen relationalen Transformation beschrieben, mittels Fremdschlüsselattributen auf eine Relation der Zeit-Dimension je nach angegebener Granularitätsstufe modelliert. Der Primärschlüssel setzt sich aus dem Fremdschlüsselattribut der Relation der Klassifikationsstufe niedriger Granularität und dem Attribut der identifizierenden zeitlichen Gültigkeit zusammen. Der Vorteil dieser Variante liegt in der Abbildung der zeitlichen Gültigkeit direkt zur Klassifikationsbeziehung. So werden die niedrige Klassifikationsstufe und damit deren beschreibende Attribute von der zeitlichen Gültigkeit der Klassifikationsbeziehung getrennt, wodurch Redundanz verringert wird. Prinzipiell handelt es sich bei diesem Konstrukt um eine 2-dimensionale „factless fact table“ (KIMBALL et al., 1998, S. 212), also eine 2-dimensionale Faktbeziehung ohne Kennzahl.

Die relationale Umsetzung des Eigenschaftsfensters für Kennzahlen stellt im Gegensatz zu allen bisher vorgestellten E/RM- als auch mE/RM-Transformationen eine Besonderheit dar. Das Eigenschaftsfenster beschreibt die Kennzahl als Modellelement und nicht die Kennzahlausprägung als Datum. Damit sind die Kennzahleigenschaften Metadaten und entsprechend als Attribute im Metamodell abzubilden (Abbildung 58). Je nach Gestaltung des Metamodells sind die Eigenschaften als Attribute einer eigenständigen Kennzahlrelation oder allgemein als Erweiterung einer Attributrelation im Data Dictionary zur Verwaltung des Datenschemas vorzusehen.

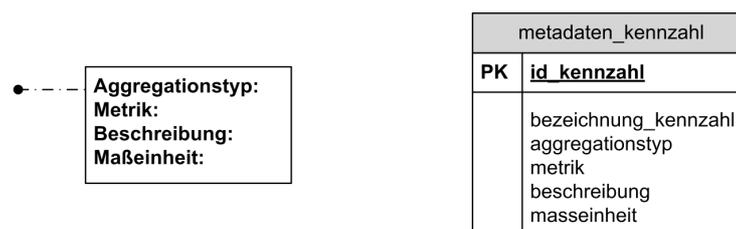


Abbildung 58: Relationale Transformation eines Eigenschaftsfensters für Kennzahlen

6.3.3 Ausgewählte Ergebnisse der relationalen Modellierung analytischer Daten

Nach Vorstellung der Transformationsregeln für die mE/RM-Notationselemente und der Erweiterungen können nun im letzten Schritt auch die semantischen multidimensionalen Teilschemata in relationale Strukturen überführt werden.

Die Dimension „Organisationseinheiten“

Die Charakteristika der Dimension „Organisationseinheiten“ im semantischen Entwurf bestanden in vorhandenen parallelen Dimensionspfaden und der zeitlich beschränkten Gültigkeit in der Ausprägung einer Reihe von Klassifikationsbeziehungen.

Die Klassifikationsstufen der Dimension werden als Relationen abgebildet, die als Primärschlüsselattribut eine ID erhalten. Wie im operativen Entwurf ist die ID des Tieres im Sachkontext durch die Lebensohrmarke (LOM) bereits eingeführt und wird entsprechend verwendet. Einfache Klassifikationsbeziehungen zur Darstellung einer zeitunabhängigen funktionalen Abhängigkeit werden wie 1:n-Beziehungen des E/RM durch einen Fremdschlüssel auf den Primärschlüssel der Relation der nächst höheren Klassifikationsstufe berücksichtigt. So enthält die Relation „euterviertel“ das obligatorische Attribut „lom“ mit einer Fremdschlüsselbeziehung auf die „lom“ der Relation „tier“ (Abbildung 59). Klassifikationsbeziehungen mit einer temporalen Restriktion als zeitabhängige funktionale Abhängigkeiten können wie beschrieben in zwei verschiedenen Varianten eine relationale Berücksichtigung finden. Hier wird der Weg einer eigenständigen Relation für die Klassifikationsbeziehung als „elegantere“ Variante gewählt. Die relationale Umsetzung gleicht damit der eines zeitgestempelten Beziehungstyps des E/RM. Die Relation für die zeitlich begrenzte Zugehörigkeit von Tieren zu Tiergruppen beispielsweise enthält Attribute für die LOM des Tieres und die ID der Tiergruppe. Beide stellen Fremdschlüssel auf die entsprechenden Relationen der Klassifikationsstufen dar. Die zeitliche Gültigkeit wird über zwei Zeit-Attribute für die Intervallgrenzen der Gültigkeit abgebildet, die wiederum Fremdschlüsselattribute auf die Klassifikationsstufe niedrigster Granularität der Zeit, also den Zeitpunkt darstellen. Die gleichen Strukturen zeigen sich auch bei den anderen Klassifikationsbeziehungen mit temporaler Restriktion in der Dimension „Organisationseinheit“ (Abbildung 59 und Abbildung 60). Die Klassifikationsbeziehungen des Teilpfades Stallabteil → Stall → Betriebsstätte weisen keine temporale Restriktion auf und werden wie bei der Klassifikationsbeziehung Euterviertel → Tier umgesetzt. Gleiches gilt für die Beziehung Abrechnungseinheit → Unternehmen.

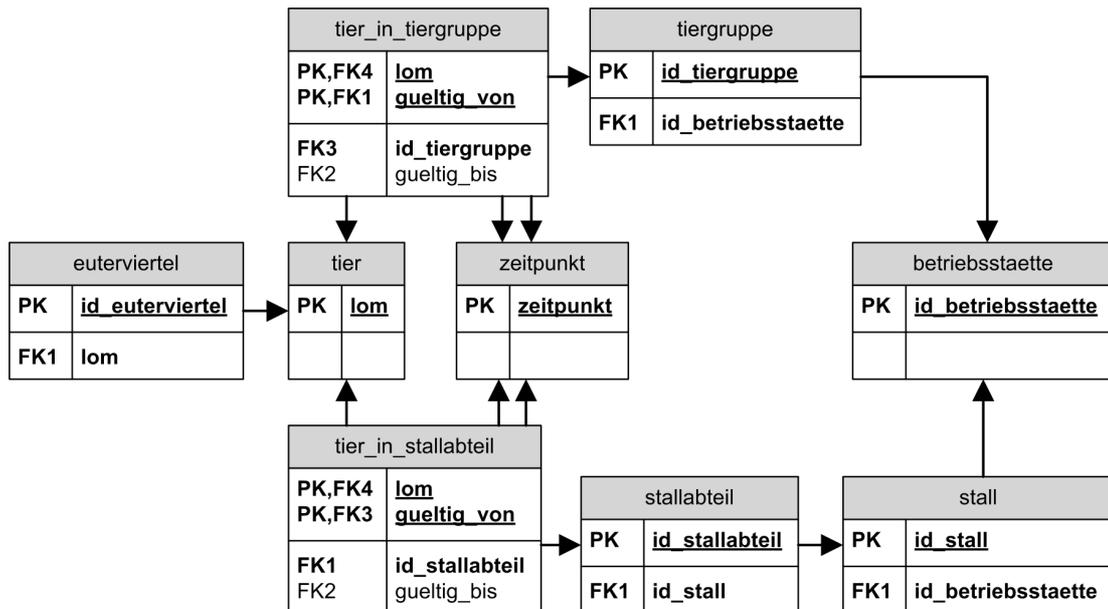


Abbildung 59: Relationales analyseorientiertes Teilschema der Dimension „Organisationseinheiten“ (Klassifikationsstufen Euterviertel, Tier, Tiergruppe, Stallabteil, Stall, Betriebsstätte)

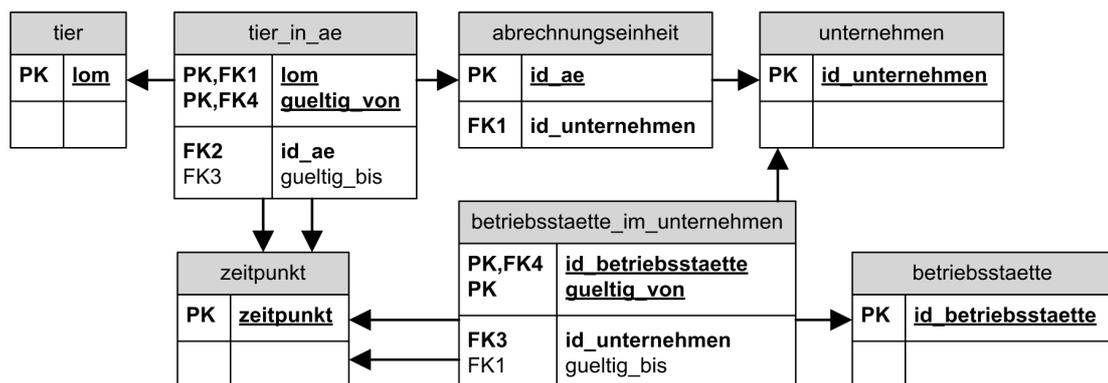


Abbildung 60: Relationales analyseorientiertes Teilschema der Dimension „Organisationseinheiten“ (Klassifikationsstufen Tier, Abrechnungseinheit, Betriebsstätte, Unternehmen)

Die Dimension „Zeit“

Die Überführung der Dimension „Zeit“ in relationale Strukturen kann unkompliziert nach den vorgestellten Transformationsregeln erfolgen. Im Unterschied zur Organisationseinheiten-Dimension tritt hier jedoch der Fall von parallelen Klassifikationspfaden unabhängig von Klassifikationsbeziehungen mit temporalen Restriktionen auf (Abbildung 61). Dabei erhält die Klassifikationsstufe, an der sich der Klassifikationspfad verzweigt, für jeden Pfad ein Fremdschlüsselattribut der Relation der nächst höheren Klassifikationsstufe. Die Relation der Klassifikationsstufe „tag“ ist aus diesem Grund mit den Fremdschlüsselattributen „id_monat“ und „id_wochentag“ modelliert, die die Relationen der Klassifikationsstufen nächst höherer Granularität der jeweiligen Parallelpfade referenzieren. Ein

Beispiel für die Ausprägung des Klassifikationspfades Zeitpunkt→Tag→Wochentag→Woche zeigt Abbildung 62.

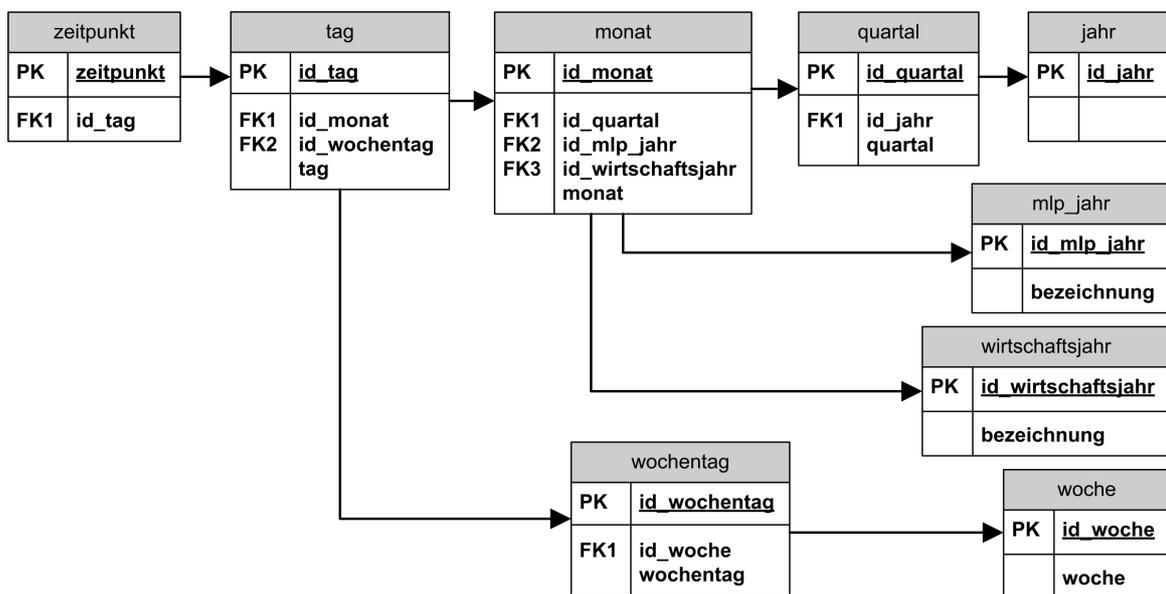


Abbildung 61: Relationales analyseorientiertes Schema der Dimension „Zeit“

zeitpunkt	<u>zeitpunkt</u>	id_tag
	2006-10-26 12:34:00	324

tag	<u>id_tag</u>	id_monat	id_wochentag	tag
	324	74	938	26

wochentag	<u>id_wochentag</u>	id_woche	wochentag
	938	134	Donnerstag

woche	<u>id_woche</u>	woche
	134	43

Abbildung 62: Beispiel für die Ausprägung des Teilpfades Zeitpunkt→Tag→Wochentag→Woche der Dimension „Zeit“

Krankheitsbehandlungen

Am Beispiel der Modellierung von Krankheitsdaten werden die vorgestellten Dimensionen zusammengeführt (Abbildung 63). Zentrales Element ist die Relation „behandlung“ als relationale Entsprechung der Faktbeziehung des mE/RM. Die zugehörigen Kennzahlen „anzahl_behandlungen“ und „anzahl_erstbehandlungen“ werden der Relation als Attribute hinzugefügt. Die Primärschlüsselattribute der Relationen für die Klassifikationsstufen niedrigster Granularität der beteiligten Dimensionen werden als Fremdschlüsselattribute modelliert. Sie bilden gleichzeitig den Primärschlüssel aus („id_tag“, „id_diagnose“,

„lom“). Auf die Umsetzung des Eigenschaftsfensters für Kennzahlen wird an dieser Stelle mit Verweis auf die Ausgestaltung des Metamodells in Kapitel 6.5.2 verzichtet.

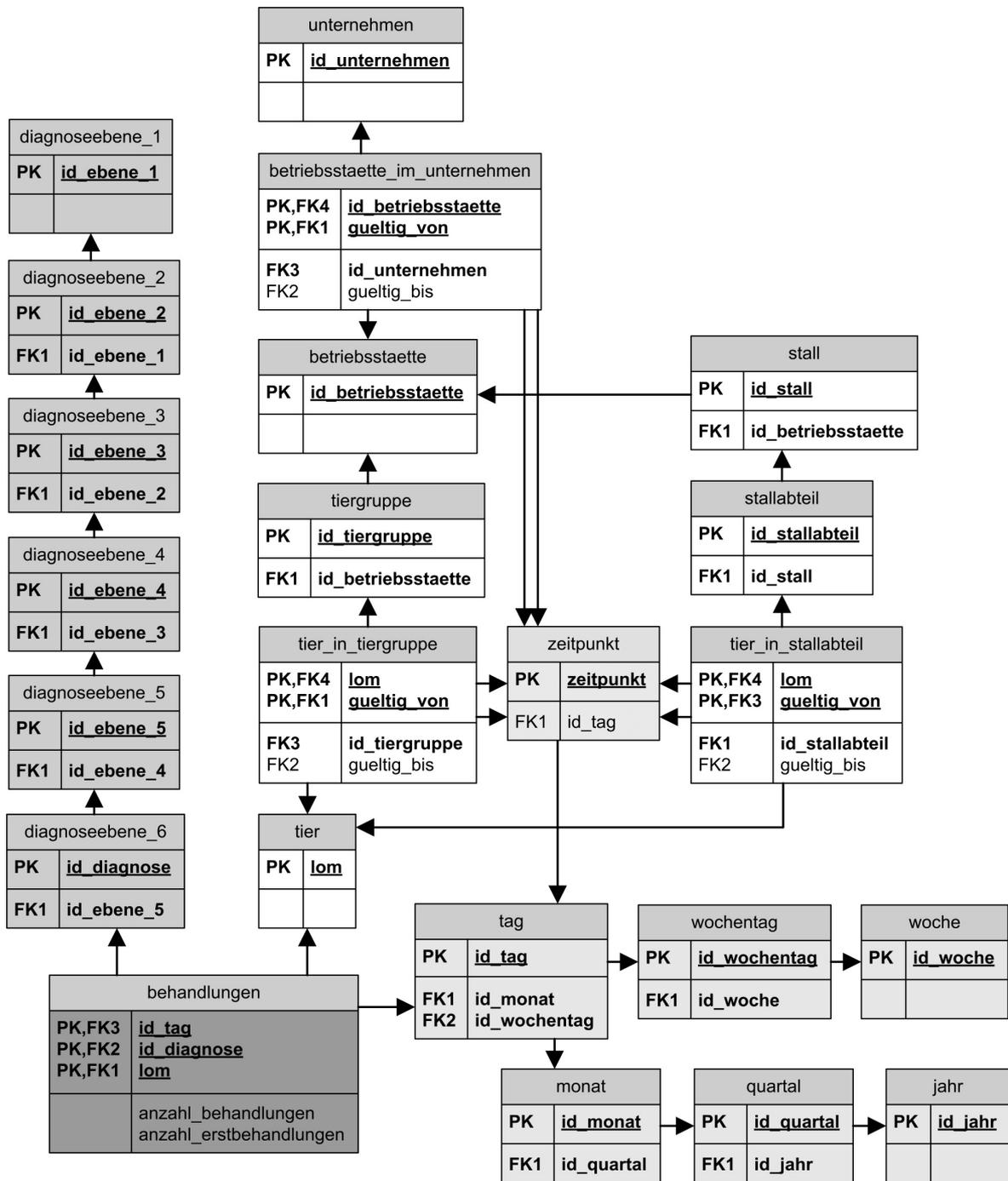


Abbildung 63: Relationales analyseorientiertes Schema für Krankheitsbehandlungen

Über die Dimension „Organisationseinheiten“, konkret durch die Modellierung der Klassifikationsbeziehungen mit temporalen Restriktionen, findet die Relation „zeitpunkt“ als Relation mit niedrigster Zeit-Granularität Eingang ins Schema, obwohl sie nicht zur Modellierung der Faktbeziehung herangezogen wird. Die 6-Ebenen-Dimension „Diagnose“ ist entsprechend den Transformationsregeln über sechs Relationen abgebildet.

Damit repräsentiert das Schema zur analyseorientierten Abbildung von Behandlungsdaten mit einzelnen Relationen für jede Klassifikationsstufe ein so genanntes Snowflake-Schema. Die im Zusammenhang der relationalen Modellierung analyseorientierter Daten oftmals als Idealform angeführte Modellierungsvariante des Star-Schemas (z.B. KIMBALL et al., 1998, S. 144 ff.) stellt dabei nur eine Möglichkeit der physischen Repräsentation dar. Sie ist darüber hinaus als Konzept denormalisierter Dimensionsschemata eher der Modellierung spezifischer Analysesichten als der Modellierung einer Basisdatenbank zuzuordnen.

6.4 Schemaintegration operativer und analytischer Daten

Die relationalen Schemata für operative und analytische Daten zeigen eine Reihe Gemeinsamkeiten. Diese sollen als Grundlage für die Zusammenführung beider relationalen Schemata als Umsetzung des hybriden Modellierungsansatzes dienen. Anhand identischer und verschiedener Schemastrukturen wird die Methodik der Schemazusammenführung im Folgenden dargestellt.

6.4.1 Identische Schemastrukturen

Der relationale Entwurf durch Anwendung von Transformationsregeln auf E/R- und mE/R-Modelle lässt identische Schemastrukturen in beiden Entwürfen erkennen. Als Beispiel werden die Modellierung der Zugehörigkeiten von Tieren zu Tiergruppen und Stallabteilen und diese wiederum zu Betriebsstätten herangezogen. Ausgehend vom E/RM (Abbildung 25 und Abbildung 26) ergibt der relationale Entwurf ein Schema mit Relationen der Entitätstypen und Relationen für die Abbildung zeitbezogener Zugehörigkeiten (Abbildung 45 und Abbildung 46). Die Umsetzung des mE/RM (Abbildung 36) führt zu einer identischen Schemastruktur (Abbildung 59), insbesondere durch die Berücksichtigung temporaler Restriktionen in der Ausprägung von Klassifikationsbeziehungen. Unterschiede ergeben sich jedoch teilweise durch die Existenz zusätzlicher Attribute in den operativ motivierten Relationen (Abbildung 64). Diese zusätzlichen Attribute sind in gemeinsame

Relationen integrierbar, so dass beide Datensichten durch entsprechende Relationen im gemeinsamen relationalen Datenmodell abgebildet werden können.

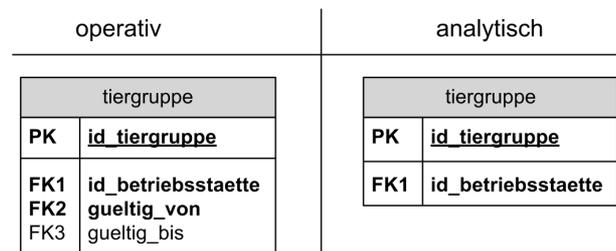


Abbildung 64: Gegenüberstellung der operativ motivierten und der analyseorientiert motivierten Relation „tiergruppe“

6.4.2 Verschiedene Schemastrukturen

Neben identischen Strukturen in beiden Teilschemata existieren jedoch auch Unterschiede. Unter der Prämisse, dass analytische Daten eine Teilmenge oder eine Aggregation operativer Daten darstellen, müssen durch analyseorientierte Schemastrukturen abgebildete Daten jedoch auch im operativen Schema eine Entsprechung haben.

In Abbildung 65 sind die Relationen des operativen und den analytischen Datenschemas zur Abbildung der Zeit gegenübergestellt. Während das operative Schema lediglich eine Relation mit einem Zeit-Attribut (Datentyp „Timestamp“) zur Abbildung von Zeitpunkten enthält, ist das analyseorientierte Schema durch mehrere Relationen zur expliziten Modellierung von Zeitpunkt-Semantik gekennzeichnet. Das heißt, ausgehend von der Abbildung des Zeitpunktes, der im analytischen Schema ebenfalls durch eine Relation repräsentiert wird, existieren Relationen für die explizite Darstellung des Tages, des Monats etc. Es handelt sich also um eine Erweiterung der operativen Datensicht. Aufgrund der Abbildung des Zeitpunktes mithilfe des Datentyps „Timestamp“ handelt es sich dabei jedoch um redundante Daten, da sie jederzeit aus einem „Timestamp“-Attribut extrahiert werden können.

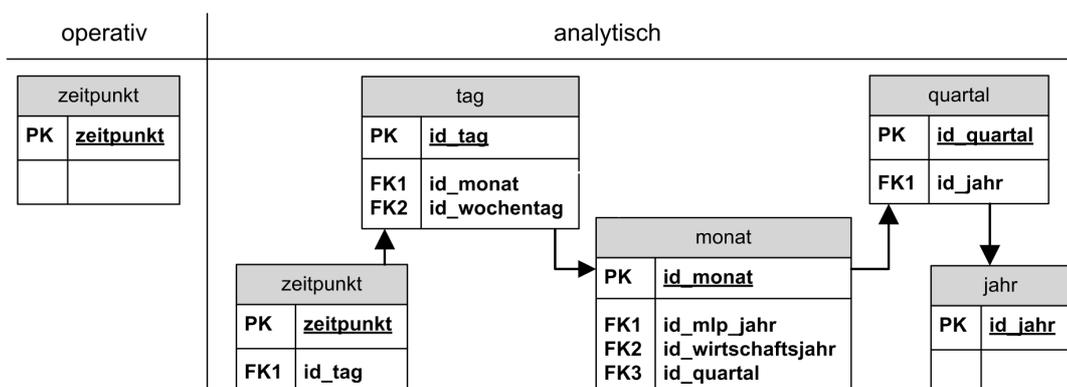


Abbildung 65: Gegenüberstellung operativer und analyseorientierter Modellstrukturen der Zeit-Dimension

Ein ähnliches Bild ergibt sich beim Vergleich der relationalen Teilschemata für Krankheitsdiagnosen (Abbildung 66). Im Unterschied zur Zeit-Dimension sind die operativen Strukturen jedoch nicht Teil des analyseorientierten Schemas. Vielmehr bieten die analyseorientierten Relationen eine redundante Sicht auf operative Daten. Die über eigenständige Relationen für jede Diagnoseebene modellierte Hierarchie des Diagnosesystems ist aus den operativ gehaltenen Daten über das Fremdschlüsselattribut „uebergeordnete_diagnose“ und das Attribut „diagnoseebene“ ableitbar.

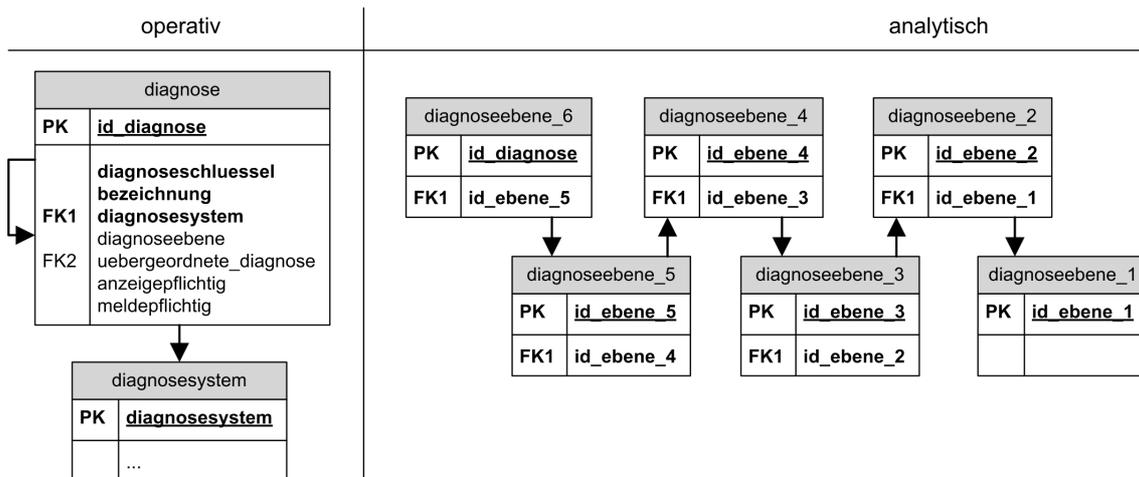


Abbildung 66: Gegenüberstellung operativer und analyseorientierter Schemastrukturen für Diagnosen

Am deutlichsten wird der aggregierende Charakter analyseorientierter Strukturen bei der Modellierung von Faktbeziehungen. Die Abbildung von Krankheitsbehandlungen an einem Tier erfolgt im operativen Schema über die Relationen „massnahme“ und „massnahme_tier“. Die Diagnose ist dagegen über die Relation „vorfall_untersuchung“ modelliert. Die Kennzahlen „anzahl_behandlungen“ und „anzahl_erstbehandlungen“ zur Beurteilung der Krankheitssituation bei Tieren, die entsprechend in der Faktbeziehungen „behandlungen“ des analyseorientierten Schemas modelliert sind, müssen folglich aus den drei genannten Relationen abgeleitet werden (Abbildung 67).

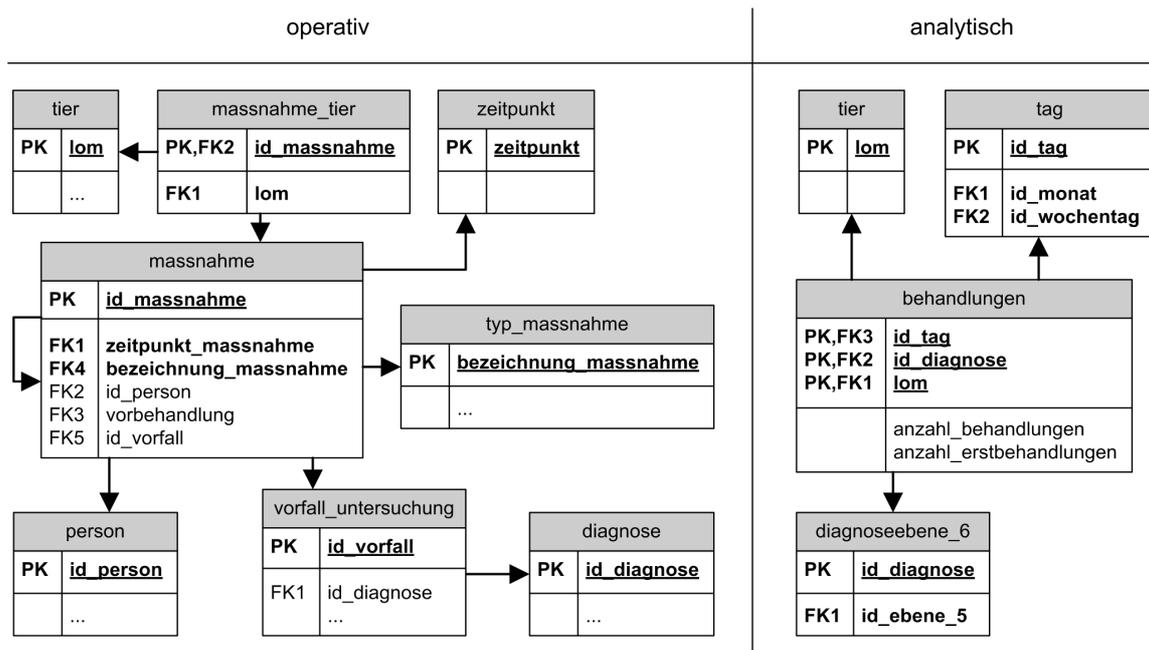


Abbildung 67: Gegenüberstellung operativer und analyseorientierter Schemastrukturen für Krankheitsbehandlungen

6.4.3 Schemazusammenführung am Beispiel von Krankheitsdaten

Am Beispiel der Krankheitsdaten wird die Zusammenführung der operativen und analyseorientierten relationalen Teilschemata zusammenfassend dargestellt (Abbildung 68).

Das gemeinsame relationale Schema kann in drei Bereiche unterteilt werden:

- operativ motivierte Schemateile
- gemeinsam genutzte Schemateile
- analytisch motivierte Schemateile

Operativ motivierte Relationen resultieren aus der Modellierung operativer Daten. Eine Einschränkung betrifft solche Daten des operativen Schemas, die weitergehende Semantik abbilden, beispielsweise die Attribute „uebergeordnete_diagnose“ und „diagnoseebene“ zur Darstellung von Hierarchien in Diagnosesystemen. Diese Information ist nicht primär im operativen Kontext notwendig, sondern im Hinblick auf eine analyseorientierte Sichtweise auf Diagnosen modelliert.

Gemeinsam genutzte Schemateile sind dagegen in gleicher Weise Gegenstand sowohl der operativen als auch der analyseorientierten Modellierung. Im Beispiel betrifft das die Relation „zeitpunkt“ und die Relationen zur Abbildung der Organisationseinheiten.

Die Gruppe der analytisch motivierten Schemateile ist als Erweiterung des operativen Schemas zu sehen. Entsprechend gekennzeichnete Relationen existieren ausschließlich zur Unterstützung analytischer Aufgaben.

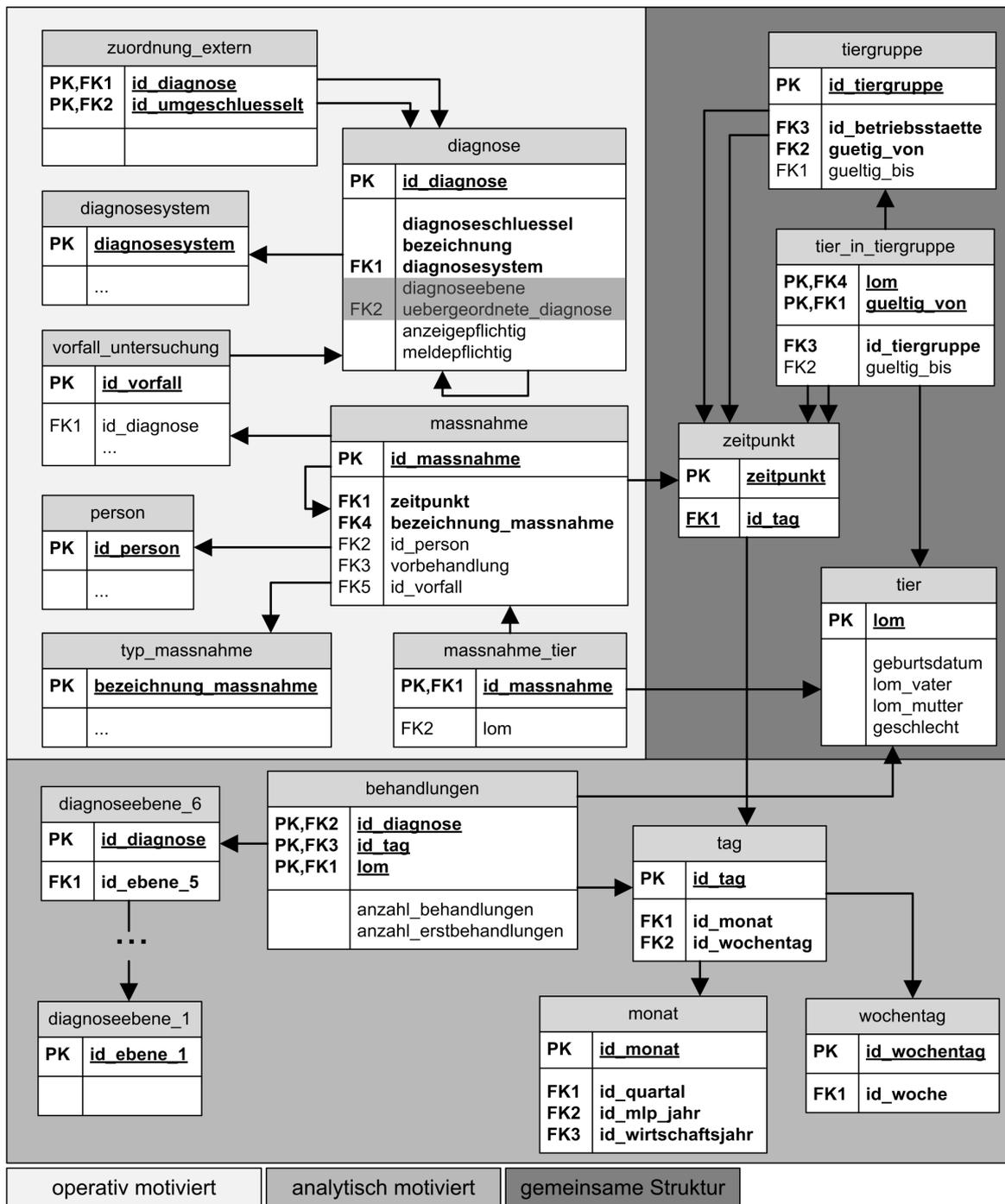


Abbildung 68: Gemeinsames relationales Schema (Ausschnitt) für Krankheitsdaten

6.5 Metamodell

Im Zuge der Beschreibung von Data Warehouse-Architekturen (Kapitel 2.2.2.2) wurde bereits auf die zentrale Rolle von Metadaten in Data Warehouse-Systemen hingewiesen. Ohne auf die Metadaten im Zusammenhang mit Datenflüssen und Verarbeitungsschritten im Data Warehouse-Prozess im Speziellen weiter einzugehen (Abbildung 7), ist es an dieser Stelle erforderlich, die Abbildung aus dem Modellierungsprozess resultierender Metadaten darzustellen. Für eine weitergehende Betrachtung wird auf das Common Warehouse Metamodel verwiesen (OMG, 2006; POOLE et al., 2003).

6.5.1 Anforderungen an das Metamodell

Die Überführung der semantischen Schemata in einen relationalen Entwurf bedeutet eine weitergehende Abstrahierung des Realitätsausschnittes. Besonders ins Gewicht fällt dies bei der Transformation des multidimensionalen Entity-Relationship-Modells. Anders als im E/RM wird im mE/RM neben Entitäts- und Beziehungstypen und Attributen analyseorientierte Semantik in Form von Klassifikationsstufen, Klassifikations- und Faktbeziehungen abgebildet. Eine Berücksichtigung entsprechender Semantik in relationalen Strukturen ist nicht adäquat möglich.

Die angeführten Beispiele zur Modellierung von Dimensionen und Faktbeziehungen zeigen die Abbildung der Datenstrukturen in Relationen. Die Rolle jener Relationen im Schema, beispielsweise als Klassifikationsstufe oder Faktbeziehung, ist jedoch nicht ersichtlich. Des Weiteren wurden mit dem Eigenschaftsfenster für Kennzahlen Elementbeschreibende Informationen in das semantische Modell aufgenommen. Auch diese sind nicht im relationalen Datenmodell darstellbar. Daraus erschließt sich die Notwendigkeit eines Metamodells für analyseorientierte Semantik, in dem beschreibende Informationen zu Modellelementen abzubilden sind.

Zusammengefasst werden können die Anforderungen an dieses Metamodell mit der Abbildung von

- Klassifikationsstufen,
- Klassifikationsbeziehungen,
- Faktbeziehungen,
- Kennzahlen und
- Eigenschaften der Kennzahlen.

Daneben besteht die Notwendigkeit der Anbindung analyseorientierter Metadaten an die Metadaten des relationalen Datenschemas. Abbildung 69 verdeutlicht den Zusammenhang

am Beispiel des zeitabhängigen Klassifikationspfades Tier→Tiergruppe. Die relationale Umsetzung des mE/RM führt zu den dargestellten Relationen mit entsprechenden Attributen. Einzig aus dem relationalen Schema ist analyseorientierte Semantik jedoch nicht mehr reproduzierbar, das heißt den relationalen Strukturen ist ihre Rolle im multidimensionalen Kontext nicht anzusehen. Für die Umsetzung multidimensionaler Auswertekonzepte ist es jedoch erforderlich, Auswertestrukturen ihrem jeweiligen Äquivalent im relationalen Datenschema zuzuordnen. Diese Aufgabe fällt einem parallel zum Datenschema zu implementierenden Metamodell zu.

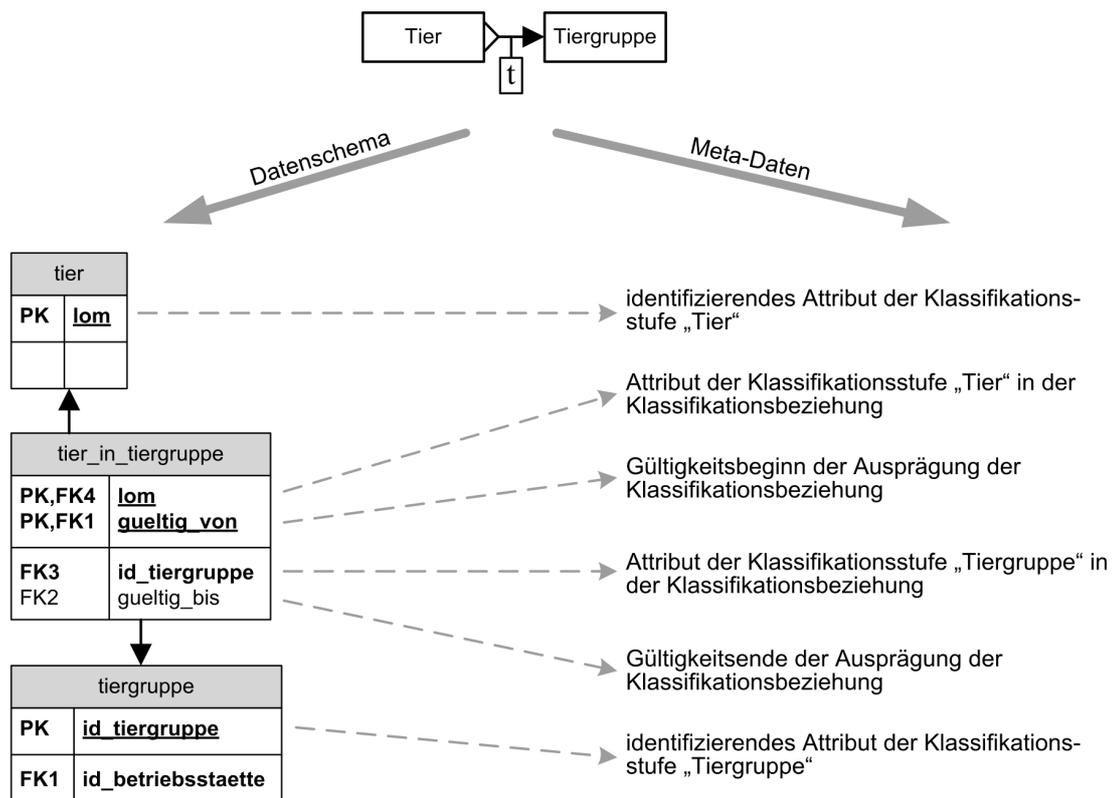


Abbildung 69: Datenschema und multidimensionale Meta-Daten

Ein E/RM des Metamodells, das die Anforderungen erfüllt, ist in Abbildung 70 dargestellt. Klassifikationsbeziehungen stellen einen Beziehungstyp zwischen Klassifikationsstufen dar und sind entsprechend modelliert. Faktbeziehungen werden zwischen n Klassifikationsstufen ausgebildet, wobei die Zuordnung als n-ärer-Beziehungstyp zwischen den Entitätstypen „Klassifikationsstufe“ und „Faktbeziehung“ umgesetzt ist. Eine Kennzahl ist eindeutig einer Faktbeziehung zugewiesen und deshalb über einen 1:n-Beziehungstyp mit dem Entitätstyp „Faktbeziehung“ verbunden.

Zur Kopplung des Metamodells für analyseorientierte Semantik an die Metadaten des relationalen Datenschemas müssen die analyseorientierten Schemaelemente in Beziehung zu Attributen des relationalen Datenschemas gesetzt werden. Aus diesem Grund ist ein Entitätstyp „relationales Attribut“ modelliert, das die Beziehung zwischen Meta- und Schemaebene herstellt. Anhand der relationalen Umsetzung des Metamodells wird dieser Aspekt weiter erläutert.

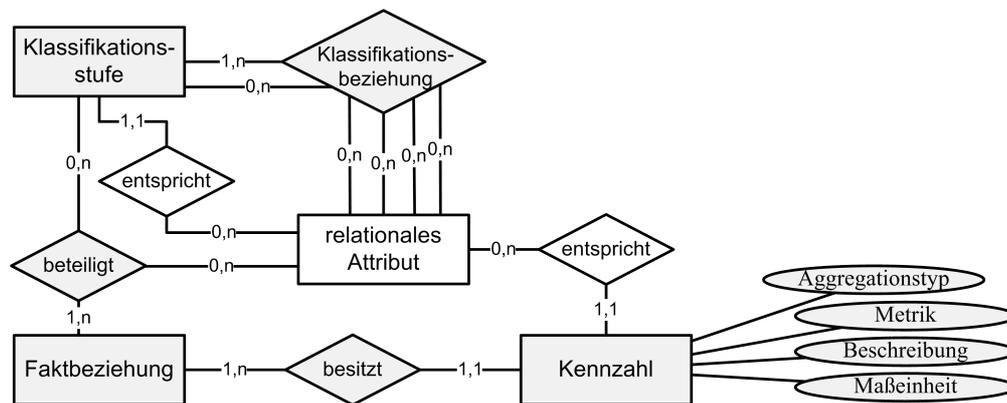


Abbildung 70: E/RM des Metamodells zur Abbildung analyseorientierter Semantik

6.5.2 Ausgestaltung des Metamodells

Die relationale Umsetzung des E/RM für das Metamodell ergibt das Schema in Abbildung 71. Als Teil der Beschreibung des relationalen Datenschemas (Data Dictionary des Datenbankmanagementsystems) ist die Relation „relationales_attribut“ dargestellt. Über diese Relation erfolgt die Kopplung zwischen dem Metamodell zur Abbildung analyseorientierter Semantik und dem Metamodell des relationalen Datenbankschemas.

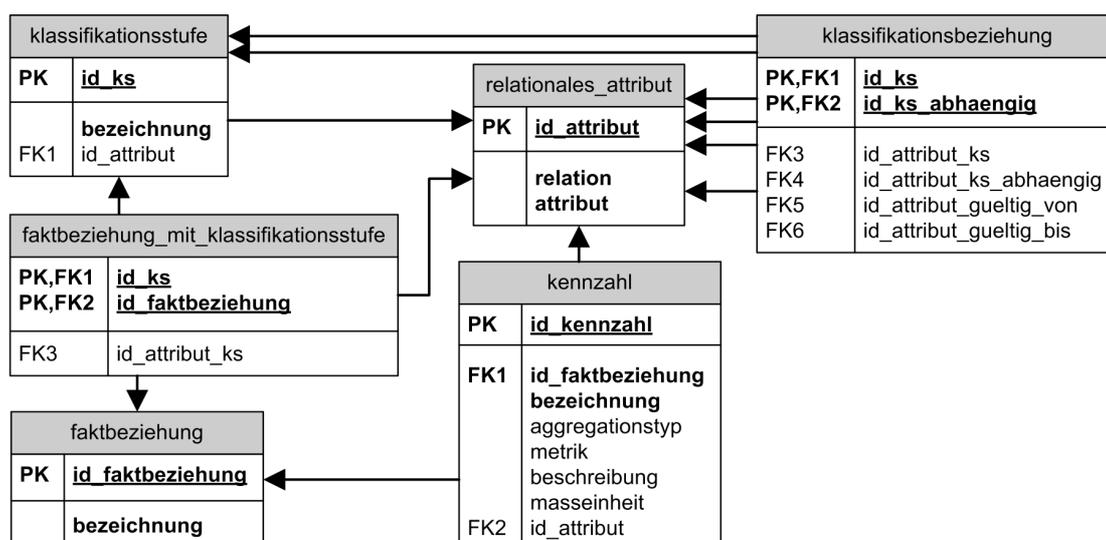


Abbildung 71: Relationales Schema des Metamodells zur Abbildung analyseorientierter Semantik

So referenziert die „id_attribut“ in der Relation „klassifikationsstufe“ ein Attribut im Datenschema, das eine Klassifikationsstufe identifiziert. Die Relation „klassifikationsbeziehung“ ist durch vier Fremdschlüsselbeziehungen zur Relation „relationales_attribut“ gekennzeichnet, um sowohl die Attribute für die beteiligten Klassifikationsstufen („id_attribut_ks“ und „id_attribut_ks_abhängig“) als auch die Attribute der Gültigkeitszeiten in der Ausprägung der Klassifikationsbeziehung („id_attribut_gueltig_von“ und „id_attribut_gueltig_bis“) angeben zu können. In adäquater Weise werden Fremdschlüsselbeziehungen zur Relation „relationales_attribut“ bei den Relationen „kennzahl“ und „faktbeziehung_mit_klassifikationsstufe“ genutzt. Um die Zusammenhänge zu verdeutlichen, ist in Abbildung 72 eine Ausprägung des Metamodells für das eingangs angeführte Beispiel der Klassifikationsbeziehung Tier→Tiergruppe (Abbildung 69) dargestellt. Im Unterschied zu Abbildung 69 ist die abzubildende multidimensionale Semantik hier jedoch über die Tabellen des erarbeiteten Metamodells dargestellt. Stellvertretend sind die Tabellen „klassifikationsstufe“ und „klassifikationsbeziehung“ abgebildet. Die Tabelle „klassifikationsstufe“ enthält Datensätze für die Klassifikationsstufen „Tier“ und „Tiergruppe“. Über ihre jeweiligen ID werden beide in der Tabelle „klassifikationsbeziehung“ zur Abbildung der temporalen Klassifikationsbeziehung in Bezug gesetzt. Damit ist der analyseorientierte Kontext im Metamodell hinterlegt. Die Nutzung entsprechender Metadaten, z.B. in auswertungsorientierter Anwendungssoftware, zwingt jedoch zur Kopplung der analyseorientierten Metadaten an das Data Dictionary des relationalen Datenschemas. Stellvertretend für das Data Dictionary ist die Tabelle „relationales_attribut“ zur Abbildung der Attribute des Datenschemas dargestellt. Durch die Referenzierung der Attribute des Datenschemas in den Tabellen „klassifikationsstufe“ und „klassifikationsbeziehung“ wird die Bedeutung der Attribute im multidimensionalen Kontext bzw. das relationale Äquivalent multidimensionaler Strukturen ersichtlich.

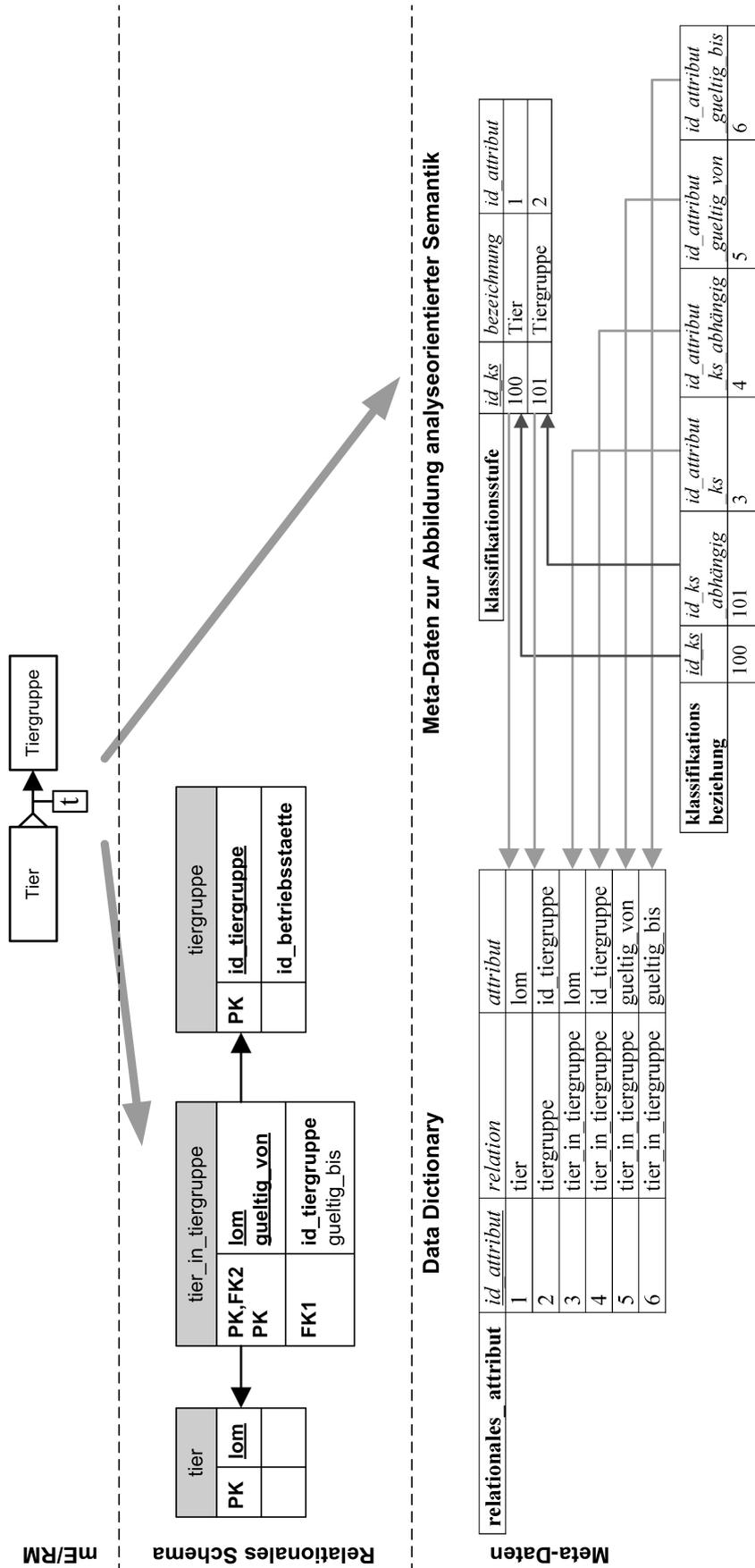


Abbildung 72: Beispiel für die Ausprägung des Metamodells

6.6 Diskussion des logischen Datenbankentwurfes

Mithilfe standardisierter Transformationsregeln wurden die semantischen Datenbankentwürfe sowohl für operative (E/RM) als auch analytische Daten (mE/RM) in relationale Schemata überführt. Gleiches gilt für die Erweiterungen des mE/RM in Bezug auf Kennzahl Eigenschaften und temporale Restriktionen in der Ausprägung von Klassifikationsbeziehungen. Als erstes Ergebnis des logischen Entwurfsschrittes liegen damit zwei getrennte relationale Datenschemata für beide Datensichten vor. Ziel des hybriden Modellierungsansatzes ist jedoch die Zusammenführung beider Teilschemata zu einem gemeinsamen relationalen Datenschema.

Beide Teilschemata sind durch eine Reihe von Gemeinsamkeiten gekennzeichnet. Das ist vor allem darin begründet, dass Klassifikationsstufen des mE/RM wie Entitätstypen, Klassifikationsbeziehungen wie 1:n-Beziehungstypen und Faktbeziehungen wie n-äre-Beziehungstypen behandelt werden. Diese Gemeinsamkeiten bilden die Grundlage der Schema-zusammenführung. Unterschiede ergeben sich dort, wo die analyseorientierte Sichtweise die atomare und originäre Abbildung des operativen Modells erweitert. Die Relationen des kombinierten relationalen Schemas können entsprechend in primär operativ motivierte, analytisch motivierte und gemeinsam verwendete Schemastrukturen klassifiziert werden. Dabei ist zu beachten, dass eine eindeutige Zuordnung nicht immer gegeben ist, weil selbst der operative Entwurf unter der Prämisse eines parallelen analyseorientierten Entwurfes entstand und eine Beeinflussung damit nicht ausgeschlossen werden kann.

Die Identifikation von gemeinsam genutzten bzw. aufeinander aufbauenden Schemastrukturen für operative und analyseorientierte Daten zeigt das Potential eines hybriden Modellierungsansatzes. Unter Ausnutzung entsprechender Effekte kann Redundanz auf Schemaebene verringert werden. Ebenso werden die Integritätsbedingungen des operativen Entwurfes bei gemeinsamen Schemastrukturen direkt und bei aufeinander aufbauenden Schemastrukturen indirekt im analyseorientierten Entwurf wirksam.

Im Zuge der physischen Implementierung des Datenschemas muss darüber entschieden werden, inwieweit Daten gemeinsamer Schemastrukturen auch gemeinsam physisch zu speichern sind oder eine Trennung operativer und analytischer Daten auf physischer Ebene angeraten ist. Ebenso ist zu beurteilen, ob analytisch motivierte Schemastrukturen als analyseorientierte Sicht auf operative Daten in Sichten oder materialisierten Sichten

implementiert werden. Damit zeigt sich erst in der physischen Implementierung, inwieweit die im hybriden Modellierungsansatz begründete verringerte Redundanz auf Schemaebene in eine verringerte Redundanz auf Datenebene umgesetzt werden kann.

Die Transformation multidimensionaler Semantik des erweiterten mE/RM in das semantikarme Relationenmodell erfordert darüber hinaus die parallele Implementierung eines Metamodells, um einen Informationsverlust zu vermeiden. Einen Mehrwert über die ausschließliche Beschreibung der multidimensionalen Strukturen hinaus liefert das Metamodell für analyseorientierte Semantik jedoch nur, wenn es als Erweiterung des Data Dictionaries zur Beschreibung des relationalen Datenschemas gesehen und entsprechend an dieses gekoppelt wird.

7 Aspekte des physischen Datenbankentwurfes

Obwohl die physische Implementierung explizit nicht Gegenstand des hybriden Modellierungsansatzes ist (Kapitel 2.3), sollen im Folgenden einige Aspekte des physischen Datenbankentwurfes zur Vervollständigung des Modellierungsprozesses vorgestellt werden. Das betrifft zum einen Aussagen zur Konsistenzsicherung und der Datenmodifikation, insbesondere vor dem Hintergrund der Speicherung gültigkeitsbezogener Daten. Zum anderen wird die Anwendbarkeit einiger Modellierungskonstrukte maßgeblich von der praktischen Ausgestaltung in einem Datenbankmanagementsystem beeinflusst. Gegenstand dieses Kapitels ist deshalb ebenfalls der Umgang mit unbestimmten Intervallgrenzen zeitlicher Gültigkeiten und mit anwendungsbezogenen Sichten auf komplexe Tabellenzusammenhänge. Ziel ist die Vereinfachung der Anwendungslogik für den Nutzer und infolgedessen die Verringerung der Fehlerträchtigkeit im Betrieb. Abgeschlossen wird das Kapitel mit einigen Aussagen zur Definition von Indizes zur Leistungsverbesserung.

Obwohl der physische Datenbankentwurf stark vom verwendeten Datenbankmanagementsystem abhängt, sollen im Folgenden hauptsächlich allgemeingültige Aussagen getroffen werden. Wo dies nicht möglich ist, wird auf die Erfahrungen bei der Implementierung einer Beispielapplikation im freien Datenbankmanagementsystem PostgreSQL in den Versionen 8.0 bis 8.2 zurück gegriffen.

7.1 Aspekte der Konsistenzsicherung

Neben den Relationsschemata gehören zur Definition eines relationalen Datenbankschemas Integritätsbedingungen zur Sicherung der Konsistenz des Datenbestandes. Der Begriff der Datenkonsistenz wird dabei als Widerspruchsfreiheit von Daten gegen entsprechend definierte Bedingungen innerhalb einer Datenbank verstanden (ZEHNDER, 2005, S. 243). LENZ (1997, S. 34) unterscheidet dabei in physische und logische Datenkonsistenz. Die Korrektheit der physischen Speicherung (Zugriffspfade, Speicherstrukturen, etc.) soll an dieser Stelle keine Rolle spielen. Vielmehr interessieren Aspekte der semantischen Konsistenz und der Datenmodellkonsistenz als Teil der logischen Datenkonsistenz. Unter dem Begriff der semantischen Konsistenz ist die Korrektheit des Datenschemas als korrekte

Abbildung des Realitätsausschnittes zu nennen, die über Integritätsbedingungen sicher zu stellen ist. Die Datenmodellkonsistenz bezeichnet die Konsistenzbedingungen, die in der Struktur des Datenmodells begründet sind (TESCHKE, 1999, S. 103). Zur Sicherung der Konsistenz in relationalen Datenbanken stehen verschiedene Mechanismen zu Verfügung (MELTON und SIMON, 1997, S. 199 ff.; ELMASRI und NAVATHE, 2002, S.236 ff.; SCHUBERT, 2004, S. 165 ff.):

- **Entitätsintegrität:**

Die Entitätsintegrität bezeichnet die Bedingung, dass keine Primärschlüsselspalte einer Tabelle den Wert NULL annehmen sowie die Eindeutigkeit des Primärschlüssels verletzen darf (SCHUBERT, 2004, S. 76). Die Entitätsintegrität ist über die Definition des Primärschlüssels Bestandteil des relationalen Datenmodells.

- **Referentielle Integrität:**

Die referentielle Integrität erfordert, dass zu jedem Wert einer Fremdschlüsselspalte einer Tabelle, der nicht NULL ist, genau dieser Wert für die referenzierte Primärschlüsselspalte der referenzierten Tabelle existiert. Diese Integritätsbedingung ist ebenfalls über die Definition eines Fremdschlüssels im relationalen Modell verankert.

- **Wertebereichseinschränkungen:**

Einschränkungen des Wertebereichs grenzen die Menge der möglichen Werte einer Spalte ein und sind über die Spaltendefinition ebenso Bestandteil des relationalen Datenmodells. Ihre Umsetzung in Datenbankmanagementsystemen erfolgt mithilfe verschiedener Mechanismen auf Basis von SQL, die zum Teil auch für die Kontrolle erweiterter Integritätsbedingungen genutzt werden (z.B. CHECK-Constraints).

- **Erweiterte Integritätsbedingungen:**

Neben den vorgestellten modellbedingten Integritätsbedingungen existieren weitere, datenbankmanagementsystemabhängige Mechanismen zur Konsistenzsicherung. Zu unterscheiden sind dabei spalten- bzw. tabellengebundene Bedingungen (CHECK-Constraint) und eigenständige Elemente relationaler Datenbankschemata wie Zusicherungen (Assertions) und Trigger (MELTON und SIMON, 1997, S. 211 ff, 387 ff.; ELMASRI und NAVATHE, 2002, S. 798).

Dem gegenüber stehen eine Reihe von Konsistenzanforderungen, die im Anwendungsfall begründet und durch entsprechende Mechanismen zu kontrollieren sind:

[a] Wertebereichseinschränkungen von Spalten (z.B. Milcheinweißgehalt < 6%)

- [b] Kontrolle von Abhängigkeiten zwischen Spaltenwerten
 - innerhalb einer Tabelle (z.B. gültig_von < gültig_bis)
 - zwischen Tabellen (z.B. Melken nur bei weiblichen, nicht trockenstellten, abgekalbten Rindern)
- [c] Sicherung einer gültigkeitsbezogenen Tupeleindeutigkeit (z.B. Eindeutigkeit von Tiergruppenversionen, Abbildung 73)
- [d] Sicherung einer gültigkeitsbezogenen referentiellen Integrität (z.B. Referenzierung von Tiergruppenversionen, Abbildung 74)
- [e] systemzeitbezogene Konsistenzsicherung (z.B. Trächtigkeitsdauer < 350 d)
- [f] Kontrolle schemabedingter Redundanz (z.B. Angabe von Vater und Mutter trotz Besamungs- und Kalbedaten)
- [g] Kontrolle entwurfsbedingter Schemasemantik (z.B. Abbildung von Vererbungssemantik im relationalen Schema)
 - Abbildung der Spezialisierungs-/Generalisierungsart (total/partiell, disjunkt/überlappend)
 - Abbildung differenzierter Eigenschaften von Sub-Entitätstypen

Zur Kontrolle von Wertebereichseinschränkung [a], von Abhängigkeiten innerhalb einer Relation [b] und von differenzierten Eigenschaften von Sub-Entitätstypen als Spalten in einer Universaltabelle [g] bietet sich die Anwendung von CHECK-Constraints. Werden vom Datenbankmanagementsystem Unterabfragen in CHECK-Constraints unterstützt, erweitern sich die Einsatzmöglichkeiten um die Kontrolle von Abhängigkeiten von Attributwerten zwischen Tupeln und zwischen Relationen [b]. Das verwendete PostgreSQL unterstützt diese Funktion jedoch nicht, weshalb auf andere Mechanismen, in diesem Fall Trigger, zurückgegriffen werden muss.

Wie bereits beschrieben, sichern die Entitätsintegrität die Eindeutigkeit des Primärschlüssels und damit eines Tupel und die referentielle Integrität die korrekte Referenzierung von Attributwerten als definierte Eigenschaften des relationalen Datenmodells. Im Zusammenhang der Abbildung gültigkeitsbezogener Daten im vorliegenden Anwendungsfall erfahren beide Mechanismen ([c], [d]) Einschränkungen in ihrer Anwendungsmöglichkeit, die an folgendem Beispiel erläutert werden sollen:

Eine Tiergruppe ist in einer gleichlautenden Tabelle modelliert. Aufgrund der Zeitstempelung zur Darstellung einer zeitlichen Gültigkeit kommt einem Tupel in der Tabelle die

Bedeutung einer Objektversion, also einer Tiergruppen-Version zu (vgl. Kapitel 5.2.3). Die Eindeutigkeit des Primärschlüssels sichert jedoch nur die formale Eindeutigkeit eines Tupel, nicht jedoch die sachliche Eindeutigkeit einer Tiergruppenversion, da diese durch die zeitlichen Gültigkeitsattribute definiert werden und zeitliche Überlappungen zwischen Versionen eines Objektes nicht ausgeschlossen werden können (Abbildung 73). Auch alternative Primärschlüssel oder UNIQUE-Constraints¹⁵ können zeitliche Überlappungen von Objektversionen nicht ausschließen (z.B. *bezeichnung* + *gueltig_von* in Abbildung 73). An dieser Stelle ist der Einsatz erweiterter Mechanismen zur Konsistenzsicherung wie Trigger, Zusicherungen oder CHECK-Constraints mit Unterabfragen nötig.

tiergruppe	<i>id_tiergruppe</i>	<i>bezeichnung</i>	<i>gueltig_von</i>	<i>gueltig_bis</i>	<i>beschreibung</i>
	1	Frischmelker	01.01.2001	31.12.2006	5. – 100. Laktationstag
	2	Frischmelker	20.10.2006	∞	5. – 120. Laktationstag

Abbildung 73: Sachlich inkonsistente Tabelle durch Zeitstempelung trotz Wahrung der Entitätsintegrität

Gleiches gilt bei Nutzung der referentiellen Integrität für Objektversionen in zeitgestempelten Tabellen:

Für die Zuordnung von Tieren zu Tiergruppen in einer eigenständigen Tabelle „tier_in_tiergruppe“ referenziert die Spalte „id_tiergruppe“ über eine Fremdschlüsselbeziehung die Primärschlüsselspalte „id_tiergruppe“ der Tabelle „tiergruppe“ (Abbildung 74). In der „klassischen“ Modellierung ohne Berücksichtigung vergangenheitsbezogener Daten enthält die Tabelle „tiergruppe“ nur Tupel aktuell gültiger Tiergruppen. Eine Referenzierung der Tiergruppe in der Zuordnung von Tieren zu Tiergruppen sichert deshalb dort die sachlich korrekte Zuordnung einer gültigen Tiergruppe über die referentielle Integrität.

tiergruppe	<i>id_tiergruppe</i>	<i>bezeichnung</i>	<i>gueltig_von</i>	<i>gueltig_bis</i>	<i>beschreibung</i>
	1	Frischmelker	01.01.2001	31.12.2006	5. – 100. Laktationstag
	2	Frischmelker	01.01.2007	∞	5. – 120. Laktationstag
	3	Krankengruppe	01.01.2001	∞	Kühe mit Milchsperr

tier_in_tiergruppe	<i>id_tier</i>	<i>id_tiergruppe</i>	<i>gueltig_von</i>	<i>gueltig_bis</i>
	31	1	03.05.2006	06.06.2006
	31	3	07.06.2006	10.06.2006
	31	2	11.06.2006	06.08.2006

Abbildung 74: Sachlich inkonsistenter Tabellenzustand durch Zeitstempelung trotz Anwendung der referentiellen Integrität

¹⁵ UNIQUE-Constraints sind ein Mechanismus zur Sicherung der Eindeutigkeit der Werte bzw. der Wertkombination einer oder mehrerer Spalten einer Tabelle unabhängig von der Primärschlüsseldefinition.

Durch die Versionisierung enthält die Tabelle „tiergruppe“ für jede Tiergruppe nicht mehr nur ein Tupel, sondern eine Vielzahl an Tupel, jedes für eine andere zeitlich gültige Version einer Tiergruppe. Eine Referenzierung der Tiergruppe durch die Tabelle „tier_in_tiergruppe“ unter Erfüllung der referentiellen Integrität stellt also nur noch sicher, dass ein entsprechendes Tupel als Wertkombination in der Tabelle „tiergruppe“ existent ist, nicht jedoch die sachliche Korrektheit im Sinne der Referenzierung einer zeitlich gültigen Tiergruppen-Version. Im Rahmen einer gültigkeitsbasierten Modellierung verliert also die referentielle Integrität weitgehend ihre Anwendungsmöglichkeit zur sachlichen Konsistenzsicherung. Es besteht also generell die Anforderung, dass die Gültigkeit der Tupel der referenzierenden Tabelle innerhalb der Gültigkeit der referenzierten Tupel liegt. Auch dies ist nur über erweiterte Mechanismen der Konsistenzsicherung zu gewährleisten.

Die Kontrolle systemzeitbezogener Bedingungen [e] unterscheidet sich von den bisher vorgestellten Bedingungen in der Form, dass nicht nur Tabellenzustände zu vergleichen sind, sondern diese in Bezug zur Systemzeit des Datenbankdienstes gesetzt werden müssen. Ein Beispiel hierfür wäre die kontinuierliche Prüfung auf sachlich nicht plausible Trächtigkeitszeiten von Kühen bei Vergleich der aktuellen Systemzeit mit dem erwarteten Kalbetermin (bei Berücksichtigung von Tierabgängen, Kalbungen, Embryo-Abgängen etc.). Hintergrund dieser Konsistenzbedingungen ist die Gefahr inkonsistenter Datenbankzustände durch fehlende Buchungen (z.B. Kalbung, Abgang). Entsprechende Konsistenzkontrollen sind losgelöst von Datenmodifikationsoperationen auf einer Datenbank auszuführen, weshalb der Einsatz von Triggern ausscheidet und auf extern aufgerufene (z.B. zeitgesteuert) Prozeduren (Stored Procedures) zurück gegriffen werden muss.

Eine weitere Art von Konsistenzbedingungen besteht in der Kontrolle schemabedingter Redundanz [f]. Obwohl Datenschemata redundanzarm zu entwerfen sind, lässt sich Redundanz nicht immer vermeiden. Im vorliegenden Datenschema tritt schemabedingte Redundanz beispielsweise in der Speicherung der Daten der Elterntiere auf. Die Tabelle „tier“ enthält die Spalten „lom_vater“ und „lom_mutter“ zur Angabe der Lebensohrmarken der Tiereltern (Abbildung 43). Die Eltern im Unternehmen geborener Tiere sind parallel dazu jedoch über das abkalbende Rind und den Sperma-liefernden Bullen ableitbar. Der Informationspfad über Besamungen → (Embryo-Transfer) → Kalbungen ergibt sich zwangsläufig durch Abbildung der entsprechenden Ereignisse. Die Angabe der Tiereltern in der „tier“-Tabelle ist daneben jedoch für dem Unternehmen zugewandene Tiere zwingend notwendig, für die nicht nachträglich Besamungen, Embryo-Transfers und

Kalbrungen ausschließlich zur Speicherung der Vorfahren-Informationen angelegt werden können. Die damit einhergehende Redundanz im Schema ist mit entsprechenden Mechanismen, z.B. CHECK mit Unterabfragen, Trigger, Assertion, zu kontrollieren.

Im Kontext der vorgestellten Modellierungsergebnisse ergeben sich auch aus der Umsetzung einer Spezialisierung/Generalisierung einzuhaltende Konsistenzbedingungen [g]. Die Spezialisierungen von technischen Objekten und Maßnahmen an Organisationseinheiten wurden als total und disjunkt eingeführt, das heißt jedes Objekt des Super-Entitätstyps ist Element nur eines Sub-Entitätstyps. Durch die relationale Umsetzung in Form von Relationen für den Super- und die Sub-Entitätstypen geht die Semantik der totalen disjunkten Teilnahme verloren. Sie ist auch mithilfe der datenmodellbedingten Integritäts-sicherungsmechanismen Entitätsintegrität, referentieller Integrität und Wertebereichs-einschränkungen nicht abbildbar. Deshalb müssen auch hier erweiterte Konsistenzsicherungsmechanismen Anwendung finden, die zum einen sicherstellen, dass ein Wert des Primärschlüsselattributes der Relation des Super-Entitätstypen genau einem Wert des Primärschlüsselattributes in einer Relation für die Sub-Entitätstypen entspricht (totale Teilnahme, beispielsweise Abbildung 48). Zum anderen muss garantiert werden, dass ein Wert des Primärschlüsselattributes der Relation eines Sub-Entitätstypen nicht gleichzeitig einem Wert des Primärschlüsselattributes der Relation eines anderen Sub-Entitätstypen entspricht (disjunkte Teilnahme).

Zusammenfassend erfordern der Anwendungsfall des Precision Dairy Farming und das vorgelegte Datenschema die Einhaltung verschiedenster Konsistenzbedingungen. Insbesondere die Speicherung gültigkeitsbezogener Daten bedarf erhöhter Aufwendungen in der Konsistenzsicherung. Denen stehen modellbasierte Kontrollmechanismen in nicht ausreichendem Maße gegenüber, so dass erweiterte, SQL-Dialekt- bzw. datenbankmanagementsystemabhängige Mechanismen zur Anwendung kommen müssen.

7.2 Aspekte der Datenmodifikation

Wie im vorhergehenden Abschnitt am Beispiel von Entitäts- und referentieller Integrität dargelegt, verändert die gültigkeitsbezogene Datenspeicherung über eine Tupel-Zeitstempelung in nicht temporal erweiterten relationalen Datenbankmanagementsystemen grundlegende relationale Eigenschaften. Das zeigt sich ebenso bei der Anwendung von Datenmodifikationsoperationen auf entsprechenden gültigkeitsbezogenen Datenbanken.

Für die Modifikation von Daten auf Tupelebene stellt SQL drei Operationen zur Verfügung:

- Einfügen (INSERT)
- Aktualisieren (UPDATE)
- Löschen (DELETE)

Abgeleitet aus der „klassischen“ Bedeutung eines Tupels als Objekt-Repräsentation (ELMASRI und NAVATHE, 2002, S. 226) stellen die genannten Datenmodifikationsoperationen Funktionalität zum Einfügen neuer bzw. Aktualisieren und Löschen bestehender Objekte zur Verfügung. Die im Zuge einer Historisierung von Daten veränderte Semantik eines Tupels von der Objekt- zur Objektversions-Repräsentation hat folglich Auswirkungen auf die Semantik und damit die Nutzungsart der Modifikationsoperationen. Üblicherweise werden die Operationen jedoch weiterhin auf die Objekt-Ebene bezogen (ELMASRI und NAVATHE, 2002, S. 801 ff.; MYRACH, 2005, S. 160 ff.), was insofern zu Missverständnissen führen kann, da die SQL-Operationen auf Tupel- also auf Objektversions-Ebene gelten. Im Folgenden werden die Modifikationsoperationen ebenfalls auf die Objekt-Ebene bezogen, der Vollständigkeit wegen enthält Tabelle 7 darüber hinaus die Semantik der Datenmodifikationsoperationen in Gültigkeitszeit-Datenbanken für beide Betrachtungsweisen.

Tabelle 7: Semantik der Datenmodifikationsoperationen in Gültigkeitszeit-Datenbanken in Objekt- und Objektversions-Betrachtungsweise

		Betrachtungsebene	
		Objekt-Sicht	Objektversions-Sicht
Datenmodifikationsoperation	Einfügen	Einfügen eines neuen Objektes durch Einfügen einer ersten Objektversion	Einfügen einer neuen Objektversion und eventuelles Ändern des Intervall-Endes der Gültigkeit der letzten Objektversion
	Aktualisieren	Aktualisieren eines Objektes durch Einfügen einer neuen Objektversion und eventuelles Ändern des Intervall-Endes der Gültigkeit der letzten Objektversion (Ausnahme: Fehlerkorrektur)	Nur das Ändern des Intervall-Endes der Gültigkeit der letzten Objektversion ist erlaubt, ansonsten Einfügen einer neuen Objektversion (Ausnahme: Fehlerkorrektur)
	Löschen	Löschen eines Objektes durch Beenden der Gültigkeit der letzten Objektversion (Ausnahme: Fehlerkorrektur)	Kein Löschen von Objektversionen (Ausnahme: Fehlerkorrektur)

Es wird ersichtlich, dass aufgrund der Historisierung von Daten, abgesehen von Fehlerkorrekturen, Modifikationen von Objektdaten in Gültigkeitszeit-Datenbanken immer zur Modifikation der Gültigkeiten von Objektversionen führen. Im Falle des Einfügens eines

neuen Objektes ist dabei eine erste Objektversion mit einem zeitlichen Gültigkeitsintervall-Anfang zu erstellen. Das Aktualisieren von Objektdaten und Löschen von Objekten ist, wieder abgesehen von Fehlerkorrekturen, in gültigkeitsbezogenen Datenbanken nicht erlaubt, da frühere Objektzustände Teil der zu speichernden Historie sind. Aus diesem Grund ist das Aktualisieren von Objektdaten durch das Anlegen einer neuen Objektversion und das Abschließen der letztgültigen Objektversion umzusetzen (Abbildung 75). In gleicher Weise bedeutet das Löschen eines Objektes nicht das Entfernen von Objektdaten aus der Datenbank, sondern das Beenden der Gültigkeit der letzten Objektversion.

tiergruppe	<i>id_tiergruppe</i>	<i>bezeichnung</i>	<i>gueltig_von</i>	<i>gueltig_bis</i>	<i>beschreibung</i>
	1	Frischmelker	01.01.2001	∞	5. – 100. Laktationstag



```
INSERT INTO tiergruppe (bezeichnung, gueltig_von, beschreibung) VALUES
('Frischmelker', '01.01.2007', '5.-120. Laktationstag');
```

```
UPDATE tiergruppe SET gueltig_bis = '31.12.2006' WHERE bezeichnung =
'Frischmelker' and gueltig_von = (SELECT MAX(gueltig_von) FROM
tiergruppe WHERE bezeichnung = 'Frischmelker' AND gueltig_von <
'01.01.2007');
```



tiergruppe	<i>id_tiergruppe</i>	<i>bezeichnung</i>	<i>gueltig_von</i>	<i>gueltig_bis</i>	<i>beschreibung</i>
	1	Frischmelker	01.01.2001	31.12.2006	5. – 100. Laktationstag
	2	Frischmelker	01.01.2007	∞	5. – 120. Laktationstag

Abbildung 75: Beispiel für das Aktualisieren eines Objektes in einer Relation¹⁶

Anhand der drei Modifikationsoperationen „Einfügen“, „Aktualisieren“ und „Löschen“ wird deutlich, dass die zeitbezogene Datenspeicherung ebenfalls höhere Aufwendungen in der Handhabung im Rahmen der Modifikation von Daten in entsprechenden relationalen, nicht temporalen Datenbanken erfordert. Maßgeblich beeinflusst wird die Handhabbarkeit jedoch daneben auch von der konkreten Implementierung der Gültigkeiten, wie der folgende Abschnitt zeigt.

7.3 undefinierte Intervallgrenzen zeitbezogener Gültigkeiten

Die zeitbezogene Speicherung von Daten über die Darstellung von Gültigkeiten in nicht temporalen Datenmodellen bzw. Datenbankmanagementsystemen führt zur Abbildung entsprechender Semantik durch benutzerdefinierte Zeit-Attribute (Kapitel 5.1.3.1). Bei der

¹⁶ Vorausgesetzt wird die Definition der Spalte „id_tiergruppe“ als Datentyp serial (selbstzählende Ganzzahl).

Modellierung von zeitlichen Gültigkeiten durch beide Intervallgrenzen, also sowohl durch den Intervall-Anfang als auch durch das Intervall-Ende, tritt die Frage der Behandlung unbestimmter Intervallgrenzen auf. Das heißt, es ist zu klären, wie eine zum aktuellen Zeitpunkt nicht spezifizierte Intervallgrenze in der Datenbank abzubilden ist. Die folgenden Ausführungen beschränken sich auf den weitaus häufigeren Fall des unbestimmten Intervall-Endes. Für den Intervall-Anfang gilt prinzipiell Gleiches. Auf semantischer Ebene sind dabei grundsätzlich drei Varianten denkbar (MYRACH, 2005, S. 123 ff.):

- nicht spezifiziert (NULL-Wert)
(da das Intervall-Ende nicht bekannt ist, ist eine Aussage nicht möglich)
- bis jetzt
(begründet durch die Annahme, dass ein Intervall ohne Angabe des Endes mindestens den aktuellen Zeitpunkt umfasst)
- größter möglicher Wert („unendlich“)
(begründet durch die Annahme, dass ein Intervall mit unbestimmter oberer Grenze ohne Änderung den größtmöglichen Wert annehmen kann)

Praktische Relevanz erfährt die Fragestellung unbestimmter Intervallgrenzen bei jeder Art des zeitpunktbezogenen Datenzugriffs, der dadurch gekennzeichnet ist, dass die Intervallgrenzen einer zeitlichen Gültigkeit mit einem spezifischen Zeitpunkt verglichen werden. Für die vorgestellten Varianten bedeutet das im Einzelnen:

- NULL-Wert: Da ein Vergleich mit einem NULL-Wert, also einem unbestimmten Zustand nicht möglich ist, muss bei einem Vergleich einerseits geprüft werden, ob der spezifizierte Zeitpunkt zwischen dem Intervall-Anfang und dem Intervall-Ende liegt und für den Fall eines unbestimmten Intervall-Endes, ob der spezifizierte Zeitpunkt größer als der Intervall-Anfang ist und das Intervall-Ende den NULL-Wert annimmt (Abbildung 76 b).
- „Bis jetzt“: Durch Ersetzen des NULL-Wertes mit der aktuellen Systemzeit wird eine Vergleichbarkeit hergestellt. Es muss abgefragt werden, ob der spezifizierte Zeitpunkt zwischen dem Intervall-Anfang und dem Intervall-Ende liegt. Durch das Datenbankmanagementsystem ist jedoch das unbestimmte Intervall-Ende permanent mit der aktuellen Systemzeit nachzuführen. Der Charakter des nicht spezifizierten Zeitpunktes als Intervall-Grenze geht verloren.
- Größter möglicher Wert („unendlich“): Durch Ersetzen des NULL-Wertes mit dem größtmöglichen Wert wird ebenfalls die Vergleichbarkeit hergestellt. Der Vergleich erfolgt, ob der spezifizierte Zeitpunkt zwischen dem Intervall-Anfang und dem

Intervall-Ende liegt. NULL-Werte in der Ausprägung des Intervall-Endes sind vom Datenbankmanagementsystem unverzüglich (beispielsweise mithilfe von Triggern) durch den größtmöglichen Wert (beispielsweise ‚infinity‘ in PostgreSQL) zu ersetzen.

Aufgrund der dynamischen Grenzverschiebung in Abhängigkeit von der Systemzeit beim „bis jetzt“-Konzept, sind nur die Varianten des NULL-Wertes und des größtmöglichen Wertes bei der Implementierung der Datenbank in die Auswahl zu ziehen. Die Vorzugslösung wird jedoch im Konzept des größtmöglichen Wertes aufgrund der leichteren Vergleichbarkeit der Intervallgrenzen in Abfragen gesehen. Die Verwendung des größten möglichen Wertes gestattet die direkte Vergleichbarkeit von Zeitpunkten mit den Intervallgrenzen (Abbildung 76 a), während bei einem möglichen NULL-Wert dessen Existenz als Zeichen einer unbestimmten Intervallgrenze zusätzlich abgefragt werden muss (Abbildung 76 b). Mit der Eliminierung von NULL-Werten in den Intervallgrenzen wird folglich die Anwendungslogik vereinfacht und somit die Fehlerträchtigkeit bei Abfragen verringert.

```
(a) SELECT id_tiergruppe FROM tiergruppe WHERE bezeichnung =  
      'Frischmelker' AND '01.05.2007' BETWEEN gueltig_von AND gueltig_bis;  
  
(b) SELECT id_tiergruppe FROM tiergruppe WHERE bezeichnung =  
      'Frischmelker' AND '01.05.2007' >= gueltig_von AND ('01.05.2007' <=  
      gueltig_bis OR gueltig_bis IS NULL);
```

Abbildung 76: Abfrage der am 1.5.2007 gültigen Version der Frischmelker-Tiergruppe auf die Tabelle „tiergruppe“ (Abbildung 74)

7.4 Konsolidierte Sicht über mehrere Relationen

Einen weiteren Schritt zur Verbesserung der Anwendungslogik umfangreicher Datenschemata stellen konsolidierte Sichten über mehrere Relationen dar. Die Anwendung des Konzeptes der totalen disjunkten Spezialisierung und dessen Umsetzung in Relationen jeweils für den Super- und die Sub-Entitätstypen führen zu einem übersichtlichen relationalen Datenschema verringerter Komplexität, insbesondere in Bezug auf die Anzahl von Fremdschlüsselbeziehungen (Abbildung 48 in Verbindung mit Abbildung 49). Die Abbildung eines Sachverhaltes über mehrere Relationen kann jedoch wenig anwenderfreundlich und in Abhängigkeit der Primärschlüsselauswahl und NOT-NULL-Bedingungen potentiell konsistenzgefährdend sein (Kapitel 7.1). Am Beispiel von Maßnahmen an Organisationseinheiten wird der Sachverhalt ersichtlich: Die Modellierung von Maßnahmen als Super-Entitätstyp und deren Spezialisierung für die verschiedenen Organisationseinheiten als

Sub-Entitätstypen (Abbildung 28 in Verbindung mit Abbildung 29) führt zu einem übersichtlichen relationalen Schema (Abbildung 48 und Abbildung 49). Abfragen zu Maßnahmen an bestimmten Organisationseinheiten erzwingen infolgedessen jedoch den Zugriff auf mehrere Tabellen mittels Verbundoperationen. Gleiches gilt für das Einfügen und Aktualisieren von Maßnahme-Daten.

Neben der nachträglichen Umsetzung anderer physischer Repräsentationen der Relationen des logischen Entwurfes besteht eine Möglichkeit der Erhöhung der Anwenderfreundlichkeit in der Etablierung integrierter Sichten auf mehrere Relationen. Für die Datenabfrage sind konsolidierte nutzerspezifische Sichten über bekannte VIEW-Konzepte der SQL möglich (Abbildung 77), die jedoch Datenmodifikationsoperationen (INSERT, UPDATE, DELETE) nicht unterstützen. Um dennoch die Bedienbarkeit für entsprechende Datenmodifikationsoperationen zu erhöhen, bietet sich der Einsatz parametrisierter Funktionen an, die bei einem Funktionsaufruf die Aufgliederung der Modifikationsoperation in Teiloperationen auf die einzelnen Relationen übernehmen. Die Funktion in Abbildung 78 zeigt ein Beispiel für die Definition einer entsprechenden Funktionalität und Abbildung 79 den dazu gehörenden Funktionsaufruf: Die Funktion erwartet bei einem Aufruf die Übergabe von 6 Parametern (Tiernummer, Zeitpunkt der Maßnahme, Maßnahmeart, ID der Diagnose, ID der behandelnden Person, ID der Vorbehandlung). Erster Schritt ist das Speichern der Diagnose über die Tabelle „vorfall_untersuchung“, das Auslesen der neu angelegten ID und ein entsprechender Eintrag für das Tier in die Tabelle „vorfall_tier“. Im zweiten Schritt wird die Maßnahme und deren Attribute, unter anderem die vorhergehend ausgelesene ID des Vorfalls, in die Tabelle „massnahme“ eingefügt. Abschließend wird ein Datensatz in der Tabelle „massnahme_tier“ mit der ID der Maßnahme und der LOM des Tieres erstellt. Über den Aufruf der Funktion, in der in Abbildung 79 dargestellten Struktur, werden die eben vorgestellten Schritte zum Einfügen einer Behandlung an einem Tier unabhängig vom Nutzer oder der Anwendungssoftware ausgeführt. Damit wird dem Nutzer eine anwenderspezifische Sicht geboten, die noch dazu über den definierten Zugriff auf die Tabellen des Schemas Maßnahmen der Konsistenzsicherung vereinfachen.

```
CREATE VIEW sicht_massnahmen_tier AS
  SELECT id_massnahme, lom, zeitpunkt_massnahme, bezeichnung_massnahme,
         id_diagnose, massnahme.id_person, vorbehandlung
  FROM massnahme_tier JOIN massnahme USING (id_massnahme)
     LEFT JOIN vorfall_untersuchung USING (id_vorfall);
```

Abbildung 77: Integrierte Sicht auf Maßnahmen am Tier über eine VIEW-Definition in der PostgreSQL-Notation auf die Tabellen „massnahme“, „massnahme_tier“ und „vorfall_untersuchung“ (Abbildung 48, Abbildung 49)

```

CREATE OR REPLACE FUNCTION fkt_massnahme_tier_einfuegen("varchar",
    "timestamp", "varchar", "int4", "int4", "int4")
    RETURNS void AS
$BODY$

DECLARE
    neue_id_vorfall int4;
    neue_id_massnahme int4;

-- $1 - Tiernummer
-- $2 - Zeitpunkt der Maßnahme
-- $3 - Bezeichnung der Art der Maßnahme
-- $4 - ID der Diagnose
-- $5 - ID der behandelnden Person
-- $6 - id_massnahme der Vorbehandlung

BEGIN

    INSERT INTO vorfall_untersuchung (id_vorfall, zeitpunkt, id_diagnose,
        id_person)
        VALUES (default, $2, $4, $5);
    -- Datentyp „serial“ für Attribut vorfall_untersuchung.id_vorfall
    -- Voraussetzung

    neue_id_vorfall := (SELECT MAX(id_vorfall) FROM vorfall_untersuchung);

    INSERT INTO vorfall_tier (id_vorfall, lom) VALUES (neue_id_vorfall,
        $1);

    INSERT INTO massnahme (id_massnahme, zeitpunkt_massnahme,
        bezeichnung_massnahme, id_vorfall, id_person, vorbehandlung)
        VALUES (default, $2, $3, neue_id_vorfall, $5, $6);
    -- Datentyp „serial“ für Attribut massnahme.id_massnahme Voraussetzung

    neue_id_massnahme := (SELECT MAX(id_massnahme) FROM massnahme);

    INSERT INTO massnahme_tier (id_massnahme, lom) VALUES
        (neue_id_massnahme, $1);

    RETURN;
END;

$BODY$
LANGUAGE 'plpgsql' VOLATILE;

```

Abbildung 78: Funktionsdefinition in der PostgreSQL-Notation zum Einfügen einer Maßnahme an einem Tier in die Tabellen „massnahme“, „massnahme_tier“ und „vorfall_untersuchung“ (Abbildung 48)

```

SELECT fkt_massnahme_tier_einfuegen(LOM des Tieres, Zeitpunkt der
    Maßnahme, Maßnahmenart, ID der Diagnose, ID der Person, ID der Vorbe-
    handlung);

```

Abbildung 79: Struktur des Aufrufes der Funktion „fkt_massnahme_tier_einfuegen“ in der PostgreSQL-Notation zum Einfügen einer Maßnahme an einem Tier

7.5 Aspekte der Indizierung

Nachdem in den vorangegangenen Abschnitten Aussagen zu ausgewählten Fragestellungen der Überführung des logischen Datenschemas in ein relationales Datenbankmanagementsystem getroffen wurden, wird das Kapitel des physischen Datenbankentwurfes mit einigen Anmerkungen zur Indizierung abgeschlossen. Dabei soll weder auf die theoretischen Grundlagen noch die verschiedenen Zugriffsmethoden eingegangen werden, sondern vielmehr der Fokus auf Aspekte der Attributauswahl für Indizes aufgrund der Tabellenstrukturen gerichtet werden. Für das Studium der Grundlagen wird auf entsprechende Datenbank-Handbücher verwiesen (z.B. LOCKEMANN und SCHMIDT, 1987, S. 226 ff.; BISKUP, 1995, S. 151 ff.).

Für die folgenden Betrachtungen ist die Definition eines Indizes als Strukturinformation im Sinne von Zeigern bzw. Zugriffspfaden auf Datensätze (als Sortierinformation zum schnellen Auffinden von Datensätzen) ausreichend. Das heißt, wie der Index eines Buches ermöglicht ein Index auf einer Tabelle ein beschleunigtes Auffinden zu selektierender Datensätze anhand der indizierten Spalte bzw. Spalten in Such- bzw. Verbundkriterien. Damit sind drei grundlegende Eigenschaften von Indizes angesprochen:

- [a] Indizes unterstützen selektive Anfragen auf Tabellen und Verbundoperationen zwischen Tabellen.
- [b] Indizes sind spaltengebunden und ihre Verwendung damit auf Anfragen, die diese Spalten betreffen, beschränkt.
- [c] Indizes als Strukturinformationen sind parallel zu den indizierten Daten gespeichert. Daraus resultiert ein Aktualisierungsaufwand bei Datenmodifikationsoperationen.

Die genannten Eigenschaften ermöglichen es, allgemeingültige Aussagen zur Indexgestaltung im Anwendungsfall abzuleiten. So resultiert aus dem hohen Aktualisierungsaufwand für Indizes [c] eine Geschwindigkeitsreduktion bei der Anwendung von Datenmodifikationsoperationen, die damit einem Geschwindigkeitszuwachs bei Anfragen entgegensteht, weshalb der Einsatz von Indizes im Einzelfall unter Abwägung der Effekte zu prüfen ist. Aufgrund der Spaltenabhängigkeit von Indizes [b] und ihrer Nutzung in selektiven Anfragen und Verbundoperationen [a] können aus einem erwarteten Anfrageverhalten grundlegende Indexstrukturen im Anwendungsfall abgeleitet werden.

In gültigkeitszeitbezogenen Datenschemata besteht aufgrund der Versionisierung von Objekten bei Anfragen eine wesentliche Anforderung in der gültigkeitszeitbezogenen Versionsselektion. Das heißt, bei transaktionalen Punktanfragen spielt die Bestimmung einer

Objektversion über ein Suchkriterium eine herausragende Rolle. Die einfache operative Beispielabfrage nach der Tiergruppe, in der sich ein Tier zu einem spezifischen Zeitpunkt befunden hat, soll dies verdeutlichen (Abbildung 80, vgl. Abbildung 45). Als Spalten der Auswahlkriterien und damit als potentielle Indexspalten fungieren im Beispiel

- die LOM des Tieres und
- die Intervallgrenzen der Gültigkeit.

```
SELECT id_tiergruppe
FROM tier_in_tiergruppe
WHERE lom = ,Beispieltier'
AND ,spezifischer Zeitpunkt' BETWEEN gueltig_von AND gueltig_bis;
```

Abbildung 80: SQL-Abfrage zur Bestimmung der Tiergruppen-Zugehörigkeit von Einzeltieren in der PostgreSQL-Notation

Für transaktionale und analyseorientierte Verbundoperationen sollen die potentiellen Indexspalten ebenfalls anhand eines einfachen Beispiels hergeleitet werden. Als Beispielabfrage fungiert die Zuordnung von Fütterungsdaten des Einzeltieres zu den Tiergruppen in denen sich die Tiere zum Zeitpunkt des Futterabrufes befanden, beispielsweise als Grundlage für Analysen (Abbildung 81). Dazu ist ein Verbund der Tabellen „fuetterung_einzeltier“, „tier_in_tiergruppe“ und „tiergruppe“ (vgl. Abbildung 44, Abbildung 45 und Abbildung 54) nötig.

```
SELECT *
FROM fuetterung_einzeltier
JOIN tier_in_tiergruppe
USING (lom)
JOIN tiergruppe
USING (id_tiergruppe)
WHERE fuetterung_einzeltier.soll_ist = 'ist'
AND startzeitpunkt BETWEEN tier_in_tiergruppe.gueltig_von
AND tier_in_tiergruppe.gueltig_bis;
```

Abbildung 81: SQL-Abfrage zu Fütterungsdaten von Einzeltieren und Tiergruppen in der PostgreSQL-Notation

Als Verbund- und Auswahlspalten und damit ebenfalls als potentielle Indexspalten werden

- die LOM des Tieres,
- die ID der Tiergruppe,
- die Unterscheidung als SOLL- bzw. IST-Datensatz und
- die Intervallgrenzen der Gültigkeit genutzt.

Beide Beispiele verdeutlichen, dass bei gültigkeitsbasierten Abfragen sowohl die untere als auch die obere Intervallgrenze Teil von Such- bzw. Verbundkriterien sind. Sie erscheinen

damit als Erfolg versprechende Indexspalten. Gleiches gilt für die LOM des Tieres und die ID der Tiergruppe zur Identifikation der Objekte bzw. der Objektversionen.

Für weitere Betrachtungen sind die drei in den Beispielen benutzten Tabellen in ihrer Rolle im Datenschema zu unterscheiden. Die Tabelle „tiergruppe“ bildet (versionisierte) Objekte der Realwelt ab und wird im multidimensionalen Kontext als Dimensionstabelle behandelt. Die Tabelle „fuetterung_einzeltier“ stellt dagegen ein Ereignis (das der Fütterung), also eine Faktbeziehung dar. Eine Sonderrolle übernimmt die Tabelle „tier_in_tiergruppe“, über die zeitbezogene Zuordnungen von Objekten abgebildet werden. Im multidimensionalen Kontext ist sie mit so genannten „Factless-Fact“-Tabellen vergleichbar.

Damit lassen sich für die drei Tabellen verallgemeinerungsfähige Schlussfolgerungen im Rahmen der Indizierung ziehen:

- Das Attribut „id_tiergruppe“ ist in der Tabelle „tiergruppe“ Primärschlüsselattribut. Ein Index auf diesem Attribut ist vor allem für Verbundoperationen hilfreich. Das deckt sich mit Aussagen von LEHNER (2003, S. 331), der Indizes auf Primärschlüsselattributen von Dimensionstabellen empfiehlt.
- In der Tabelle „fuetterung_einzeltier“ sind die Attribute „lom“, „startzeitpunkt“ und „soll_ist“ Fremdschlüsselattribute, denen in Suchkriterien und Verbundoperationen eine hohe Bedeutung zugemessen wird. Die Indizierung dieser Spalten wird als vorteilhaft angesehen. Auch LEHNER (2003, S. 331) schlägt die Anlage von Indizes auf den Fremdschlüsselattributen von Faktbeziehungen vor. Die Art der Indexstruktur und vor allem die Reihenfolge der Attribute bei Entscheidung für einen mehrteiligen Index sind jedoch anhand des spezifischen Anfrageverhaltens zu definieren.
- Wie die Beispiele gezeigt haben, kommt den Intervallgrenzen der zeitlichen Gültigkeit in der Zuordnungstabelle „tier_in_tiergruppe“ bei Punktanfragen und Verbundoperationen eine große Bedeutung zu. Hier bietet sich die Anlage eines mehrteiligen Indizes, bestehend aus unterer und oberer Intervallgrenze, zur Beschleunigung entsprechender Zugriffe an. Die alleinige Verwendung der unteren Intervallgrenze in Indizes, z.B. im Rahmen von Primärschlüsselindizes, wird als nicht ausreichend erachtet.

Obwohl die Implementierung von Indexstrukturen die Kenntnis des Zugriffsverhaltens auf Datenbanken erfordert, lassen sich dennoch allgemeingültige Aussagen zur Indizierung aus der Tabellenstruktur ableiten. Dabei fällt die Abhängigkeit potentieller Indexstrukturen von multidimensionalen Schemastrukturen auf, ebenso wie die bestimmende Rolle der zeitlichen Gültigkeiten.

8 Schlussfolgerungen

Die Zielsetzung der vorliegenden Arbeit wurde mit der Erstellung eines Referenzmodells als Vorschlag der Datenorganisation für den Bereich der Milcherzeugung unter Anwendung eines hybriden Modellierungsansatzes benannt. Der hybride Modellierungsansatz wurde dabei als Methodik des parallelen Datenbankentwurfes für operative und analyseorientierte Daten eingeführt und angewandt, so dass eine Bewertung im Sinne abzuleitender Schlussfolgerungen für den semantischen und den logischen Entwurf erfolgen kann.

Mit dem semantischen Datenbankentwurf erfolgt die Umsetzung des durch die Informationsbedarfsanalyse charakterisierten Realitätsausschnittes in ein formales Modell. Dabei ist die anwendungsbezogene Datensicht mithilfe geeigneter Modellierungsnotationen abzubilden. Für die Wahl eines konkreten Modellierungsansatzes ist daher die Anwendungsemantik ausschlaggebend. Bei der Darstellung von Kennzahleigenschaften und temporalen Restriktionen in der Ausprägung von Klassifikationsbeziehungen wurden dennoch Defizite in bestehenden Modellierungsansätzen am Beispiel des mE/RM deutlich, die entsprechend durch Notationserweiterungen auszugleichen sind. Der hybride Modellierungsansatz als paralleler Entwurfsprozess operativer und analytischer Daten eröffnet prinzipiell die Möglichkeit, beide Datensichten bereits im semantischen Entwurf in einem gemeinsamen Schema zusammen zu führen. Gerade die Verwendung des mE/RM als analyseorientierte Erweiterung des E/RM ermöglicht ein entsprechendes Vorgehen. Dagegen spricht jedoch die grundlegende Aufgabe des semantischen Entwurfes in einer anwendungsbezogenen Abbildung des Realitätsausschnittes. Ein gemeinsames semantisches Modell für verschiedene anwendungsbezogene Datensichten kann dieser Anforderung jedoch nicht genügen, so dass eine Schemaintegration auf dieser Entwurfsebene nicht erfolgen sollte.

Der logische Datenbankentwurf repräsentiert die Ebene des Implementierungsdatenmodells. Das heißt, im logischen Entwurf ist die anwendungsbezogene Datensicht des semantischen Schemas in Strukturen eines logischen Datenmodells zu überführen, die die Basis der physischen Datenspeicherung bilden. Damit kommt diesem Entwurfsschritt die

Aufgabe der Schemazusammenführung auf dem Weg zu einem gemeinsamen Datenschema zu. Die Zusammenführung der Schemata beider Datensichten hat dabei auf der Basis identischer und aufeinander aufbauender Schemastrukturen zu erfolgen. Potentielle Konfliktpunkte bei der Integration verschiedener Datenschemata bestehen gemeinhin in (ELMASRI und NAVATHE, 2002, S. 580 f.):

- [a] Benennungskonflikte: Verwendung verschiedener Bezeichnungen in den Schemata für identische Objekte oder Merkmale
- [b] Typkonflikte: Darstellung eines Sachverhaltes durch unterschiedliche Konstrukte, beispielsweise Modellierung als Entitätstyp oder Attribut
- [c] Verschiedene Wertemengen: Definition des gleichen Attributes über unterschiedliche Wertmengen
- [d] Verschiedene Einschränkungen: unterschiedliche Einschränkungen in den Schemata, beispielsweise bezüglich Primärschlüssel und Beziehungskardinalitäten

Sowohl Benennungskonflikte [a] als auch Konflikte aufgrund verschiedener Wertemengen [c] lassen sich durch den parallelen Entwurfsprozess operativer und analytischer Daten vermeiden. Typkonflikte [b] und Konflikte zwischen Einschränkungen [d] sind dagegen zum Teil in der Logik der anzuwendenden Transformationsregeln (Kapitel 6.2.1 bzw. 6.3.1) begründet und durch Anwendung eines hybriden Modellierungsprozesses nicht zwangsläufig vermeidbar. Es zeigt sich jedoch, dass die Anwendung der beschriebenen Modellierungsgrundsätze (Kapitel 5.2.1.3), vor allem die Nutzung von Entitätstypen für Organisationseinheiten, Typ-Entitätstypen und die explizite Zeit-Modellierung in Form von Ereignis- und Gültigkeitszeiten dazu führt, dass operative Strukturen im relationalen Datenschema denen zur Abbildung analyseorientierter Daten ähneln und damit eine Schemaintegration durch verringerte Typ- und Einschränkungskonflikte begünstigt wird. Ergebnis des logischen Entwurfsschrittes ist ein gemeinsames logisches Datenschema zur Abbildung operativer wie auch grundlegender analyseorientierter Daten. Darauf aufbauend stellt sich die Frage nach der physischen Repräsentation jener Schemastrukturen, die als erweiterte Sicht auf den operativen Datenbestand interpretiert werden können. Grundsätzlich sind mit dem vorliegenden Datenschema zwei verschiedene Strategien der Datenorganisation denkbar, in deren Folge das Datenschema den Ausgangspunkt verschiedener Datenbank-Implementierung bilden kann:

[a] Trennung operativer und analytischer Daten auf physischer Ebene in getrennten Informationssystemen mit den optionalen Datenbank-Implementierungen (Abbildung 82):

- integrierte operative Datenbasis
- Operational Data Store
- Data Warehouse Datenbank

[b] Konsolidierte Datenbasis zur gemeinsamen Organisation operativer und analyseorientierter Daten in einem Informationssystem mit grundsätzlich zwei Varianten der Speicherung analyseorientierter Daten:

- Sichtenansatz (physische Speicherung operativer Daten und Abbildung analyseorientierter Daten als SQL-Sichten)
- physische Speicherung analyseorientierter Daten im Sinne materialisierter Sichten¹⁷

Damit sind auch die Auswirkungen des hybriden Modellierungsprozesses differenziert nach der Nutzungsrichtung des Datenschemas zu betrachten.

Die Strategie [a] entspricht der üblichen Datenorganisation durch Trennung der operativen und der analytischen Datensicht in verschiedenen Informationssystemen (Kapitel 2.2). Dessen ungeachtet birgt die Anwendung des hybriden Modellierungsansatzes Vorteile für den Entwurf der notwendigen Datenbank-Infrastruktur entsprechender Konzepte, die vor allem im Bereich der schemabedingten Datenkonsistenz (LENZ, 1997, S. 33; TESCHKE, 1999, S. 99 ff.; LEHNER, 2003, S. 161 ff.) in Data Warehouse Systemen zu sehen sind (Abbildung 82):

1. Die Einführung einer integrierten operativen Datenbasis sichert eine schemabedingte Multi-Quellenkonsistenz, das heißt die horizontale Schemaintegration auf operativer Ebene sichert die Konsistenz der Quelldaten.
2. Die Nutzung des Datenschemas zur Implementierung einer Data Warehouse Basisdatenbank oder einer konsolidierten Data Warehouse Datenbasis sichert in gleicher Weise die schemabedingte Multi-Basis- bzw. die Multi-Sichtenkonsistenz.
3. Die parallele Einführung einer integrierten operativen Datenbasis und einer Data Warehouse Datenbank basierend auf dem gleichen logischen, hybrid modellierten Datenschema vereinfacht aufgrund gleicher bzw. ähnlicher Datenstrukturen der implemen-

¹⁷ Als materialisierte Sicht wird an dieser Stelle die physische Speicherung von Daten verstanden, die prinzipiell auch durch SQL-Views repräsentiert werden könnten.

tierten Datenbanken den Prozess der Datenbeschaffung, konkret die Transformation der Daten. Gleiches gilt bei paralleler Implementierung eines Operational Data Store. Dem gegenüber ist die zeitliche Komponente des Konsistenzzustandes eines Data Warehouse Systems in Folge unterschiedlicher Aktualisierungsniveaus innerhalb, aber auch zwischen den Datenverarbeitungsebenen verständlicherweise unabhängig vom verwendeten Modellierungsansatz. Vielmehr werden die zeitliche Komponente der Multi-Sichtenkonsistenz und die Basis- bzw. Sichtenkonsistenz von der zeitlichen Ausführung von Datenverarbeitungsschritten bestimmt, die den Konsistenzstatus innerhalb und zwischen den Ebenen verändern.

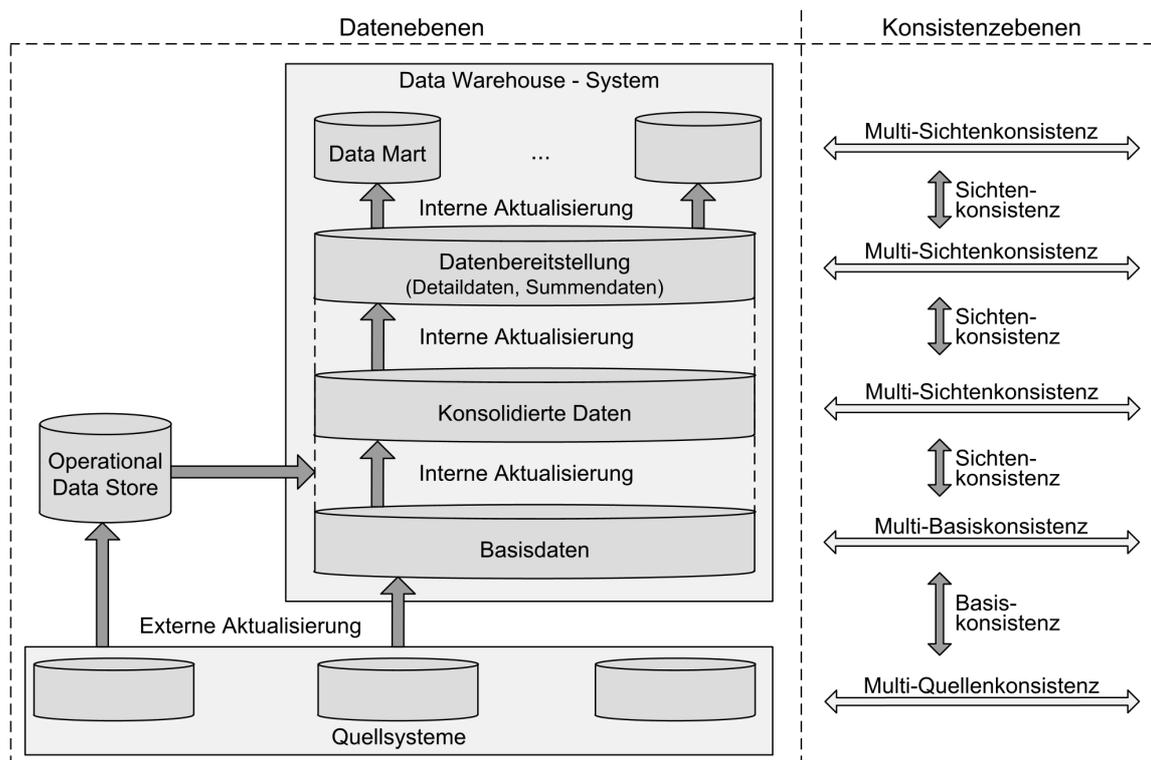


Abbildung 82: Daten- und Konsistenzebenen eines Data Warehouse-Systems (nach LEHNER, 2003, S. 23, 163)

Aufgrund der konzeptbedingten physischen Trennung operativer und analyseorientierter Daten dieser Implementierungsstrategie kann der Vorteil der reduzierten Redundanz und Inkonsistenz auf Schemaebene nicht zwangsläufig auf die Datenebene übertragen werden. Vorteile ergeben sich dennoch durch Vermeidung schemabedingter Inkonsistenz, gerade bei Implementierung von Datenbanken verschiedener Datenverarbeitungsebenen auf Basis des gleichen logischen Schemas. Der Hauptnutzen der Trennung beider Anwendungsrichtungen in unabhängigen Informationssystemen ist jedoch eindeutig im Laufzeitverhalten beider Systeme zu sehen. Die unabhängigen Systeme beeinträchtigen sich nicht

gegenseitig im Betrieb und die Datenbanksysteme können für die jeweilige Nutzungsrichtung optimiert werden.

Mehr Vorteile des hybriden Modellierungsansatzes kommen bei Umsetzung der Implementierungsstrategie einer konsolidierten Datenbasis zur gemeinsamen Organisation operativer und analyseorientierter Daten (Strategie [b]) als konsequente Fortführung der Idee der hybriden Datenmodellierung zum Tragen. Durch die Implementierung einer gemeinsamen Datenbank für die Daten beider Aufgabenfelder kann der Vorteil verringerter Redundanz und der verringerten Gefahr von Inkonsistenz auf Schemaebene direkt auf Datenebene übertragen werden. Das ist darin begründet, dass durch Implementierung gemeinsam genutzter und aufeinander aufbauender Schemastrukturen Daten zur Unterstützung beider Nutzungsrichtungen tatsächlich nur einmal gespeichert werden. Die verringerte Gefahr von Inkonsistenz beruht zum einen darauf, dass durch Vermeidung der Daten-Mehrfachspeicherung gleichzeitig die Gefahr sinkt, dass sich mehrmalig gespeicherte Daten widersprechen. Zum anderen nimmt die hybride Modellierung die zur nachträglichen Einführung einer analyseorientierten Datensicht notwendige Schemaintegration vorweg, so dass auf diesem Weg ebenfalls ein Beitrag zur Konsistenzsicherung geleistet wird. Damit verbunden ist ebenfalls der Vorteil, dass die Konsistenzbedingungen des operativen Umfeldes direkt auch für die darauf aufbauende analyseorientierte Datensicht gelten, wodurch die Qualität analyseorientierter Daten verbessert werden kann. Unterschieden werden müssen jedoch die Implementierungsvarianten des Sichtenansatzes, materialisierter Sichten und Mischformen zwischen beiden.

Der Sichtenansatz repräsentiert die Idealform der Redundanzvermeidung, indem Daten einmalig physisch gespeichert und daraus abgeleitete analyseorientierte Sichten zur Laufzeit erstellt werden. Der Nachteil ist damit jedoch klar in den ressourcenzehrenden Anfragen während der Laufzeit zu sehen. Der Ansatz materialisierter Sichten umgeht diesen Nachteil durch Vorberechnung der abgeleiteten Sichten und deren physischer Speicherung, wodurch Anfragen beschleunigt, Inkonsistenzen aufgrund unterschiedlicher Aktualisierungsniveaus dagegen begünstigt werden. Grundsätzlich bleibt jedoch der Nachteil bestehen, dass sich bei der Implementierung beider Datensichten in einer gemeinsamen Datenbank die Nutzungsrichtungen im Betrieb gegenseitig beeinflussen und nutzungsabhängige Optimierungsmaßnahmen nur in Kompromissen umgesetzt werden können.

Für den landwirtschaftlichen Anwendungsfall dürfte die Strategie einer gemeinsamen Datenbasis trotz der erwarteten Nachteile im Laufzeitverhalten eine besondere Rolle

spielen. Aufgrund der vorherrschenden kleinen landwirtschaftlichen Unternehmensstrukturen sind Investitionen in informationstechnische Infrastruktur, die eine Trennung beider Datennutzungsrichtungen zwangsläufig nach sich ziehen, zu überdenken. Des Weiteren sind landwirtschaftliche Unternehmen durch flache Organisationshierarchien gekennzeichnet, so dass der Umfang des zugriffsrelevanten Personenkreises für analyseorientierte Anfragen als gering eingeschätzt werden kann. Damit ist eine gegenseitig negative Beeinflussung der operativen und analyseorientierten Nutzungsrichtungen im Laufzeitverhalten nur im geringen Maße zu erwarten. Dazu sind jedoch entsprechende Untersuchungen anhand typischer Anwendungen in produktionsnahen Implementierungen durchzuführen.

Grundsätzlich lässt die allgemeine Schemastruktur mit der Unterscheidung in Organisationseinheiten, Merkmalstyp-Relationen, Ereignis- und Zuordnungs-Relationen sowie die explizite Zeit-Modellierung (Ereignis- und Gültigkeitszeiten) eine Übertragbarkeit der Ergebnisse auf andere Nutztierarten, andere landwirtschaftliche und darüber hinaus außerlandwirtschaftliche Anwendungsfelder erwarten.

9 Zusammenfassungen

9.1 Zusammenfassung

Das Konzept des Precision Dairy Farming als integrativer Ansatz zur nachhaltigen und effizienten Erzeugung von Milch beruht in besonderem Maße auf der Nutzung einzeltierbezogener Daten. Dabei werden hohe Anforderungen an das Datenmanagement gestellt, um erfasste Daten langfristig konsistent zu speichern, bedarfsgerecht zur Verfügung zu stellen und zu entscheidungsrelevanten Informationen aufzuwerten.

Eine Analyse des Datenmanagements in der Landwirtschaft, insbesondere der Milcherzeugung, zeigt Defizite in der Datenorganisation, die eine sinnvolle Datennutzung behindern bzw. Datenqualitätsdefizite erklären. Als Konsequenz ist eine grundlegende Neugestaltung der Datenorganisation mit dem Schwerpunkt der Datenmodellierung zu fordern. Dabei ist sowohl auf eine prozessübergreifende Sicht zur Unterstützung operativer Aufgaben als auch auf die Unterstützung analytischer Aufgaben zu achten. Die Auswertung von Konzepten der Datenorganisation in der Informatik zeigt die strikte Trennung operativer und analyseorientierter Informationssystemwelten. Diese Trennung beider Aufgabenfelder muss jedoch nicht zwangsläufig in zwei verschiedenen Datenschemata münden. Im Rahmen der Umsetzung des Precision Dairy Farming-Konzeptes bietet sich die einmalige Möglichkeit der gleichzeitigen Modellierung entsprechender Sichtweisen in einem gemeinsamen Entwurfsprozess an. Als Lösungsansatz wird deshalb ein hybrider Modellierungsansatz als paralleler Datenbankentwurfsprozess operativer und analyseorientierter Daten vorgeschlagen, in dessen Ergebnis ein gemeinsames relationales Datenschema als Basis für operative und analyseorientierte Informationssysteme vorliegt.

Ausgangspunkt des Modellierungsprozesses ist die Informationsbedarfsanalyse zur Erfassung der Anforderungen an ein zu erstellendes Informationssystem. Die durchgeführte Informationsbedarfsanalyse mit dem methodischen Schwerpunkt der Expertenbefragung zur Identifizierung zukünftiger Erfordernisse bestätigt die zentrale Rolle des Einzeltieres

im Konzept des Precision Dairy Farming und verdeutlicht die Komplexität der Milcherzeugung durch die hohe Anzahl zu berücksichtigender Merkmale.

Ausgehend von den in der Informationsbedarfsanalyse identifizierten Merkmalen, erfolgte im zweiten Schritt der semantische Datenbankentwurf. Für die Abbildung des operativen Kontextes wurde das Entity-Relationship-Model genutzt. Aufgrund unzureichender Semantik zur Darstellung analyseorientierter Zusammenhänge, fand das multidimensionale Entity-Relationship-Modell nach SAPIA et al. (1998) bei der Modellierung analyseorientierter Daten Anwendung. Auch dessen Darstellungselemente genügten nicht der vollständigen Abbildung der analyseorientierten Sichtweise des Anwendungsfalls, so dass eigene Notationselemente für die Kennzeichnung temporaler Restriktionen in der Ausprägung von Klassifikationspfaden und für Kennzahleigenschaften eingeführt werden mussten.

Das Ziel des hybriden Modellierungsansatzes, ein gemeinsames Datenschema zur Unterstützung operativer und analyseorientierter Aufgaben zu erstellen, wurde mit dem nachfolgenden relationalen Modellierungsschritt verfolgt. Ausgehend von den getrennten Datenschemata auf semantischer Ebene wurden anhand standardisierter Transformationsregeln für die Notationselemente des E/RM, des mE/RM als auch der Erweiterungen zur Kennzeichnung temporaler Restriktionen und Kennzahleigenschaften relationale Datenschemata erstellt. Das Ziel der Schemaintegration beider Datensichten erfolgte auf Grundlage der Identifikation identischer, aufeinander aufbauender und unabhängiger Schemastrukturen. Daraus ergibt sich ein modularer Aufbau des gemeinsamen Datenschemas für operative und analyseorientierte Daten.

Die Überführung höherwertiger multidimensionaler Semantik in das relationale Modell erfordert darüber hinaus die parallele Etablierung eines Metaschemas, um einen Informationsverlust in Bezug auf die analyseorientierte Sichtweise zu verhindern.

Ergebnis des hybriden Modellierungsprozesses ist damit ein relationales Datenschema, das operative wie auch analyseorientierte Daten des Anwendungsfalls beschreibt. Der spezifische Vorteil des hybriden Modellierungsansatzes wird in der engen Verzahnung beider Datensichten auf Schemaebene gesehen. Die entwurfsprozessinhärente Schemaintegration verringert die Gefahr von Redundanz und Inkonsistenz durch Schemakonflikte bei Implementierung operativer und analyseorientierter Datenbanken auf Basis des vorgestellten logischen Datenschemas. Darüber hinaus werden die operativen Konsistenzbedingungen auf die analyseorientierte Datensicht übertragen. Dies stellt einen wichtigen Schritt zur

Erhöhung der Datenqualität auswertungsorientierter Datenbanken dar. Inwieweit diese Vorteile tatsächlich auch auf die Datenebene übertragen werden können, hängt maßgeblich von der physischen Implementierung in einer konsolidierten Datenbank oder der Trennung operativer und analyseorientierter Daten auf physischer Ebene ab.

Darüber hinaus lassen die vorgelegten, verallgemeinerungsfähigen Ergebnisse eine Anwendbarkeit über den Bereich des Precision Dairy Farming hinaus erwarten.

9.2 Summary

Precision Dairy Farming, as an integrative approach to dairy farming, is especially based on the analysis of data of individual cows. This sets high standards for data management in order to store aggregated data permanently and consistently, to provide data according to demand, and to upgrade information relevant to decision making.

The examination of data management within an agricultural context, particularly dairy farming, shows deficits in data organisation that restrict the use of the data, and explain deficits regarding the quality of the data. Therefore, this paper attempts a new structuring of data organisation with the focus on data modelling to remedy the above-mentioned deficits. In order to do so, both a process-oriented perspective to support operational tasks, as well as support of analytical tasks, are necessary. The analysis of data organisation concepts in computer science shows a strict separation of operational and analysis-oriented information systems. However, this separation of tasks does not necessarily result in two different data schemes. Hence, we propose a parallel modelling procedure in a common design process. The approach taken is a hybrid modelling process integrating operational and analysis-oriented data. This results in a joint relational data scheme as a basis for operational and analysis-oriented information systems.

Our starting point is an analysis of the information requirements in order to determine the requirements of the information system. The implemented analysis of the information requirements methodologically focuses on expert interviews in order to identify future needs. This analysis confirms the key role of the individual cow within the framework of Precision Dairy Farming, and illustrates the complexity of milk production due to the large variety of respective criteria identified by the requirement analysis.

Based on these criteria, the next step is semantic data modelling. We used the Entity Relationship Model in order to map the operational context. In the course of modelling the analysis-oriented data, the multidimensional Entity Relationship Model was also applied,

due to the lack of rich semantics to map analytical relations. However, these notational elements do not sufficiently represent the application-related analysis-oriented perspective. Therefore, we introduced new notation elements for marking temporal restrictions within the values of the classification paths, and for marking the properties of measures.

The semantic modelling is followed by a relational modelling. Based on separate data schemes on a semantic level, relational data schemes were created, applying standardised transformation rules for the notational elements of E/RM, mE/RM, as well as for the extensions marking temporal restrictions and properties of measures. Our goal, the integration of both data views, was achieved by identifying identical, consecutive, and independent scheme structures. As a result, we receive a modular structure of joint data schemes of operational and analysis-oriented data.

The mapping of rich multi-dimensional semantics into the relational model requires furthermore a simultaneous introduction of a meta-scheme to avoid the loss of information with respect to an analysis-oriented point of view.

The result of this hybrid modelling process is a relational data scheme which describes both operational, as well as analysis-oriented data. However, the specific advantage of this hybrid modelling approach has to be seen in the close interconnection of both data views on a data level. The integration within the design process reduces the risk of redundancy and inconsistency caused by conflicts of schemes in the course of the implementation of the operational and analysis-oriented databases. Furthermore, the conditions of operational consistency are transferred to an analysis-oriented data view. This is an important step to increase the quality of analysis-oriented databases. The extent to which these advantages can be applied at the data level will largely depend on the subsequent physical implementation in the form of a consolidated database, or a separation of operational and analysis-oriented data on a physical level.

The given results can be expected to be applied to areas other than Precision Dairy Farming.

Literaturverzeichnis

- ALLEN, J. F. (1983): Maintaining Knowledge about Temporal Intervals. *Communications of the ACM*, Jg. 26, Nr. 11, S. 832–843.
- AMMON, C.; SPILKE, J. (2005): Comparison of fixed- and random-regression models using different functional approaches of lactation curves for milk yield forecasts. *Proceedings EFITA/WCCA 2005 in Vila Real, Portugal*. Univ. de Trás-os-Montes e Alto Douro. Vila Real, Portugal, S. 630–635.
- AMMON, C.; SPILKE, J. (2006): Probleme und Möglichkeiten der Nutzung einzeltierbezogener Daten in Milchviehbetrieben. In: Wenkel, K. O.; Wagner, P.; Morgenstern, M.; Eisermann, P. (Hrsg.): *Land- und Ernährungswirtschaft im Wandel. Referate der 26. GIL Jahrestagung, Potsdam*. LNI P-78, S. 21–24.
- ATC (Agricultural Telematics Centre) / Taurus (Hrsg.) (1993): *Design TAURUS Standard Interface Version 2.0*. Wageningen.
- AUERNHAMMER, H. (1988): Einbindung der Prozesssteuerung in das rechnergestützte Management. In: Bundesforschungsanstalt für Landwirtschaft (Hrsg.): *Elektronikeinsatz in der Tierhaltung. Landbauforschung Völkenrode. Sonderheft 93*. Braunschweig-Völkenrode, S. 311–326.
- BALZERT, H. (2005): *UML 2 kompakt*. 2. Aufl., Spektrum, München.
- BAUER, A.; GÜNZEL, H. (2004): *Data-Warehouse-Systeme. Architektur, Entwicklung, Anwendung*. 2. Aufl., dpunkt, Heidelberg.
- BAUFÖRDERUNG LANDWIRTSCHAFT GMBH (2005): *ISOagriNet*. Online verfügbar unter <http://www.isoagrinet.org>, zuletzt geprüft am 28.12.2007.
- BEIERSDORF, H. (1995): Informationsbedarf und Informationsbedarfsermittlung im Problemlösungsprozess ‚Strategische Unternehmensplanung‘. Hampp, München.
- BEERS, G. (2002): State of the art of tracking and tracing in Dutch agribusiness. In: Wild, K.; Müller, R.A.E.; Birkner, U. (Hrsg.): *Referate der 23. GIL-Jahrestagung in Dresden 2002 (Band 15)*, S. 15–19.
- BERG, E.; SCHMIDT-PAULSEN, T. (1990): Expertensysteme in der strategischen Unternehmensplanung. In: Buchholz, H.E.; Neander, E.; Schrader, H. (Hrsg.): *Technischer Fortschritt in der Landwirtschaft. Tendenzen, Auswirkungen, Beeinflussung*. Schriften der Gesellschaft für Wirtschafts- und Sozialwissenschaften des Landbaus e.V., Band 26. Landwirtschaftsverlag. Münster-Hiltrup, S. 213–221.
- BILL, R.; KORDUAN, P. (2004): Projektinformationssystem. In: Kuratorium für Technik und Bauwesen in der Landwirtschaft e. V. (Hrsg.): *Managementsystem für den ortsspezifischen Pflanzenbau. Verbundprojekt pre agro, Abschlussbericht, Kap. 6*, KTBL, Darmstadt. S. 3-72.

- BISKUP, J. (1995): Grundlagen von Informationssystemen. Vieweg & Sohn, Braunschweig.
- BLÖNNIGEN, P. (1990): Gestaltung eines Management-Informationssystems für Marktfruchtbetriebe. In: Deutsche Landwirtschafts-Gesellschaft e.V. (Hrsg.): Integrierte Programme für landwirtschaftliche Betriebe. Eine Zwischenbilanz. Arbeitsunterlagen B/90. DLG-Verlag, Frankfurt, S. 71–86.
- BODMER, U.; HORVÁTH, L. (2002): "Gläserne Produktion" von Fleisch unter Berücksichtigung von Informationstechnologien. Zeitschrift für Agrarinformatik, Jg. 10, Nr. 4, S. 54–60.
- BOEHNLEIN, M.; ULBRICH-VOM ENDE, A. (1999): Deriving initial data warehouse structures from the conceptual data models of the underlying operational information systems. DOLAP '99: Proceedings of the 2nd ACM international workshop on Data warehousing and OLAP, S. 15–21.
- BÖTTINGER, S.; OETZEL, K.; AUTERMANN, L.; BRILL, V. (2004): Software Betriebe. In: Kuratorium für Technik und Bauwesen in der Landwirtschaft e. V. (Hrsg.): Managementsystem für den ortsspezifischen Pflanzenbau. Verbundprojekt pre agro, Abschlussbericht, Kap. 4, KTBL, Darmstadt. S. 3-70.
- BRAND, A.; FOLKERTS, H.; HANEKAMP, W. J.A.; DE HOOP, W. D.; VERHEIJEN, G. M.A. (1995): Information model for dairy farms. Agricultural Telematics Centre, Wageningen.
- BULOS, D.; FORSMAN, S. (2006): Getting started with ADAPT. OLAP Database Design. White Paper, Symmetry Corporation, San Rafael, USA.
- CHAMONI, P.; GLUCHOWSKI, P. (1999a): Analytische Informationssysteme - Einordnung und Überblick. In: Chamoni, P.; Gluchowski, P. (Hrsg.): Analytische Informationssysteme. Data Warehouse, On-Line Analytical Processing, Data Mining. Springer, Berlin. S. 3–25.
- CHAMONI, P.; GLUCHOWSKI, P. (1999b): Entwicklungslinien und Architekturkonzepte des OLAP. In: Chamoni, P.; Gluchowski, P. (Hrsg.): Analytische Informationssysteme. Data Warehouse, On-Line Analytical Processing, Data Mining. Berlin. Springer, S. 261–280.
- CHEN, P. P. (1976): The entity-relationship model - toward a unified view of data. ACM Transactions on Database Systems (TODS). Vol. 1, No.1, S. 9–36.
- CODD, E. F. (1970): A Relational Model of Data for Large Shared Data Banks. Communications of the ACM, Jg. 13, Nr. 6, S. 377–387.
- CODD, E. F.; CODD, S. B.; SALLY, C. T. (1993): Providing OLAP to User-Analysts: An IT Mandate. E. F. Codd & Associates.
- DLG (Deutsche Landwirtschafts-Gesellschaft e.V.) (Hrsg.) (1985): Pflichtenheft für die Herdenführung und Fütterung in der Milchproduktion. Anforderungen an Computerprogramme. DLG-Arbeitsunterlagen A/85. DLG-Verlag, Frankfurt.
- DLG (Deutsche Landwirtschafts-Gesellschaft e.V.) (Hrsg.) (1987): Pflichtenheft für die Datenverarbeitung in der Pflanzenproduktion. Expertenwissen für Landwirte, Berater und Programmierer. Arbeitsunterlagen L/87. DLG-Verlag, Frankfurt.

- DLG (Deutsche Landwirtschafts-Gesellschaft e.V.) (Hrsg.) (1990): Integrierte Programme für landwirtschaftliche Betriebe. Eine Zwischenbilanz. Arbeitsunterlagen B/90. DLG-Verlag, Frankfurt.
- DLG (Deutsche Landwirtschafts-Gesellschaft e.V.) (Hrsg.) (1992): Integrierte Datenverarbeitung - ein Datenbankmodell für den Marktfruchtbetrieb. DLG-Arbeitsunterlagen A/92. DLG-Verlag, Frankfurt.
- DE MOL, R. M. (2000): Automated detection of oestrus and mastitis in dairy cows. Dissertation. Universität Wageningen, Wageningen.
- DEVIR, S.; RENKEMA, J. A.; HUIRNE, R. B. M.; IPEMA, A. H. (1993): A New Dairy Control and Management System in the Automatic Milking Farm. Basic Concepts and Components. *Journal of Dairy Science*, Jg. 76, S. 3607–3616.
- DINTER, B.; SAPHIA, C.; HÖFLING, G.; BLASCHKA, M. (1998): The OLAP market: state of the art and research issues. Proceedings of the 1st ACM international workshop on Data warehousing and OLAP, Washington, D.C., USA. ACM Press. New York, S. 22–27.
- DOLUSCHITZ, R. (2002): Betriebszweig- Betriebs- und Unternehmensebene der landwirtschaftlichen Produktion. In: Doluschitz, R.; Spilke, J. (Hrsg.): *Agrarinformatik*. UTB. Stuttgart, S. 266–321.
- DOLUSCHITZ, R.; BAUR, H. (1997): *Erfolgsfaktor Information*. Deutscher Sparkassenverlag, Stuttgart.
- DOLUSCHITZ, R.; KUNISCH, M. (2004): agroXML - ein standardisiertes Datenformat für den Informationsfluss entlang der Produktions- und Lieferkette. *Zeitschrift für Agrarinformatik*, Nr. 4, S. 65–67.
- ELMASRI, R.; NAVATHE, S. B. (2002): *Grundlagen von Datenbanksystemen*. 3. Aufl., Pearson Studium, München.
- ENGLER, J.; TÖLLE, K.-H.; TIMM, H. H.; HOHLS, E.; KRIETER, J. (2005): Control charts applied to individual sow farm analysis. In: Cox, S. (Hrsg.): *Tagungsband Precision Livestock Farming 2005*. Wageningen Academic Publishers. Wageningen, S. 319–325.
- FETTKE, P.; LOOS, P. (2005): Zur Identifikation von Strukturanalogien in Datenmodellen. Ein Verfahren und seine Anwendung am Beispiel des Y-CIM-Referenzmodells von Scheer. *Wirtschaftsinformatik*, Jg. 47, Nr. 2, S. 89–100.
- FISCHER, J. (1992): *Datenmanagement: Datenbanken und betriebliche Datenmodellierung*. Oldenbourg, München.
- FISCHER, R.; GROENEVELD, E.; BERGFELD, U. (2002): APIIS ein neuer Ansatz zur Erstellung konsistenter Datenbanken in der Tierzucht. In: Wild, K.; Müller, R.A.E.; Birkner, U. (Hrsg.): *Referate der 23. GIL-Jahrestagung in Dresden 2002 (Band 15)*.
- GABRIEL, R.; GLUCHOWSKI, P. (1998): Grafische Notationen für die semantische Modellierung multidimensionaler Datenstrukturen in Management Support Systemen. *Wirtschaftsinformatik*, Jg. 40, Nr. 6, S. 493–502.
- GOLFARELLI, M.; MAIO, D.; RIZZI, S. (1998a): Conceptual Design of Data Warehouses from E/R Schemes. Proceedings Hawaii International Conference on System Sciences, Kona, Hawaii. S. 334–343.

- GOLFARELLI, M.; MAIO, D.; RIZZI, S. (1998b): The Dimensional Fact Model: A Conceptual Model for Data Warehouses. *International Journal of Cooperative Information Systems*, Jg. 7, Nr. 2-3, S. 215–247.
- GRAF, R. (1990): Erschließung elektronischer Datenbankdienste für Programme im landwirtschaftlichen Betrieb. In: Deutsche Landwirtschafts-Gesellschaft e.V. (Hrsg.): *Integrierte Programme für landwirtschaftliche Betriebe. Eine Zwischenbilanz. Arbeitsunterlagen B/90*. DLG-Verlag, Frankfurt, S. 181–186.
- GROENEVELD, E. (2004): An adaptable platform independent information system in animal production: framework and generic database structure. *Livestock Production Science*, Nr. 87, S. 1–12.
- GROENEVELD, E.; YORDANOVA, L.; HIEMSTRA, S.J. (2004): Organizational structure and information technological support of national gene banks. *Livestock Production Science*, Nr. 89, S. 297–304.
- GÜNZEL, H. (2001): Darstellung von Veränderungen im multidimensionalen Datenmodell. Dissertation, Arbeitsberichte des Instituts für Informatik, Bd. 34, Nr. 8. Universität Erlangen-Nürnberg, Erlangen.
- HANNUS, T.; POIGNEE, O.; SCHIEFER, G. (2003): Implementierung eines webbasierten Werkzeugs zur Qualitätskommunikation im Rahmen eines Supply Chain weiten Qualitätsmanagement – Konzepts. In: Budde, H.-J.; Müller, R. A. E.; Birkner, U. (Hrsg.): *Referate der 24. GIL-Jahrestagung in Göttingen 2003 (Band 16)*.
- HANSEN, H. R. (1996): *Wirtschaftsinformatik 1. Grundlagen betrieblicher Informationsverarbeitung*. 7. Aufl., UTB, Stuttgart.
- HARPIN, H. H. (1990): Entwicklung eines operationalen Entscheidungs-Unterstützungssystems für landwirtschaftliche Unternehmen. Dissertation, Fachbereich Agrarwissenschaften. Justus-Liebig-Universität Gießen
- HARREN, A.; HERDEN, O. (1999): MML und mUML - Sprache und Werkzeuge zum konzeptionellen Data Warehouse-Design. In: *Proceedings 2. Workshop "Data Mining und Data Warehousing als Grundlage moderner entscheidungsunterstützender Systeme"*. LWA99 Sammelband. Universität Magdeburg, Magdeburg, S. 57–67.
- HIRSCHAUER, N. (2001): Controlling. In: Odening, M.; Bokelmann, W. (Hrsg.): *Agrarmanagement*. Stuttgart. Ulmer, S. 267–339.
- INMON, W.H. (1996): *Building the Data Warehouse*. 2. Aufl., Wiley, New York.
- INMON, W. H. (1999): *Building the Operational Data Store*. 2. Aufl., Wiley, New York.
- JENSEN, C. S.; DYRESON, C. E. (Editors) (1998): The Consensus Glossary of Temporal Database Concepts. In: Etzion, O.; Jajodia, S.; Sripada, S. (Hrsg.): *Temporal Databases - Research and Practice*. LNCS 1399. Springer, Berlin, S. 367–405.
- KAISER, A. (2000): *Die Modellierung zeitbezogener Daten*. Lang, Frankfurt am Main.
- KARSTEN, S.; STAMER, E.; JUNGE, W.; LÜPPING, W.; KALM, E. (2005): The use of a database for genetic evaluation and to manage health in dairy cows. *Proceedings of the 56th EAAP, Uppsala, Book of abstracts*, S. 381.

- KARSTEN, S.; STAMER, E. (2006): Datenerfassungssysteme für Bullenmütterprüfung und Testherden. In: Deutsche Gesellschaft für Züchtungskunde e.V. (Hrsg.): Neue Herausforderungen für die Rinderzucht. 4. Rinderworkshop, DGfZ-Schriftenreihe (Heft 43), S. 172–179.
- KEMPER, A.; EICKLER, A. (2004): Datenbanksysteme. 5. Aufl., Oldenbourg, München.
- KEMPER, H.G.; LEE, P.-L. (2003): The Customer-Centric Data Warehouse - an Architectural Approach to Meet the Challenges of Customer Orientation. Proceedings of the 36th Hawaii International Conference on System Sciences. IEEE. Washington, DC, USA.
- KIMBALL, R. (1997): Relocating the ODS: moving the operational data store will solve a number of problems. DBMS, Jg. 10, Nr. 13, S. 12–14.
- KIMBALL, R.; REEVES, L.; ROSS, M.; THORNTON, W. (1998): The Data Warehouse Lifecycle Toolkit. Tools and Techniques for Designing, Developing, and Deploying Data Warehouses. Wiley, New York.
- KLEINHENZ, B.; JÖRG, E.; KEIL, B.; KLUGE, E.; ROBBERT, D. (1998): Integrierter Pflanzenschutz - Rechnergestützte Entscheidungshilfen. Köllen, Bonn.
- KLINDT WORTH, M. (1999): Prozeßsteuerung in der Milchviehhaltung. DLG-Merkblatt 312. DLG-Verlag, Frankfurt.
- KNOLMAYER, G.; MYRACH, T. (1996): Zur Abbildung zeitbezogener Daten in betrieblichen Informationssystemen. Wirtschaftsinformatik, Jg. 38, Nr. 1, S. 63–74.
- KÖHLER, W.; SCHACHTEL, G.; VOLESKE, P. (1996): Biostatistik. Einführung in die Biometrie für Biologen und Agrarwissenschaftler. Springer, Berlin.
- KTBL (Kuratorium für Technik und Bauwesen in der Landwirtschaft e. V.) (Hrsg.) (2005): AgroXML. Online verfügbar unter <http://www.agroxml.de>, zuletzt geprüft am 28.01.2007.
- KUHLMANN, F.; KÜBLER, H.; MÜLLER, H.; WAGNER, P. (1990): Entscheidungsunterstützung für landwirtschaftliche Unternehmen durch integrierte Planungs- und Kontrollrechnungen. In: Deutsche Landwirtschafts-Gesellschaft e.V. (Hrsg.): Integrierte Programme für landwirtschaftliche Betriebe. Eine Zwischenbilanz. Arbeitsunterlagen B/90. DLG-Verlag, Frankfurt. S. 20–32.
- LASSMANN, W.; PICHT, J.; ROGGE, R. (2001): Wirtschaftsinformatik Kalender 2002. IM Marketing-Forum, Ettlingen.
- LEHNER, W. (2003): Datenbanktechnologie für Data-Warehouse-Systeme. Konzepte und Methoden. dpunkt Verlag, Heidelberg.
- LENZ, R. (1997): Adaptive Datenreplikation in verteilten Systemen. Teubner, Stuttgart.
- LENZ, H.-J.; SHOSHANI, A. (1997): Summarizability in OLAP and Statistical Data Bases. Proceedings of the 9th International Conference on statistical and scientific Database Management (SSDBM'97). Olympia, USA, S. 132–143.
- LINSEISEN, H.; SPANGLER, A.; HANK, K.; WAGNER, P.; STEINMAYR, T.; DEMMEL, M.; AUERNHAMMER, H.; MANAKOS, I.; SCHNEIDER, T.; LIEBLER, J. (2000): Daten, Datenströme und Software in einem Informationssystem zur teilflächenspezifischen Pflanzenproduktion. Zeitschrift für Agrar-informatik, Jg. 8, Nr. 2, S. 36–43.

- LINSEISEN, H. (2002): Entwicklung eines Management-Informationssystems für Entscheidungen im Precision Farming. Dissertation. Technische Universität München, Freising-Weihenstephan.
- LKV (Landeskontrollverband) Nordrhein-Westfalen e.V. (Hrsg.) (2004): ADIS-ADED Einführung. Online verfügbar unter <http://www.edi-agrar.org/>, zuletzt geprüft am 28.12.2007.
- LUDOWICY, C.; SCHWAIBERGER, R.; LEITHOLD, P. (2002): Precision Farming. Handbuch für die Praxis. Frankfurt: DLG-Verlag.
- MARTINI, D.; SPIETZ, C.; KLOEPFER, F. (2006): Darstellung teilflächenspezifischer Maßnahmen in agroXML als Voraussetzung für die Verwendung im Bereich Precision Farming. In: Wenkel, K. O.; Wagner, P.; Morgenstern, M.; Eisermann, P. (Hrsg.): Land- und Ernährungswirtschaft im Wandel. Referate der 26. GIL Jahrestagung in Potsdam. Köllen. Bonn, S. 161–164.
- MAYR, H.C.; DITTRICH, K.R.; LOCKEMANN, P.C. (1987): Datenbankentwurf. In: Lockemann, P. C.; Schmidt J.W. (Hrsg.): Datenbank-Handbuch. Berlin. Springer, S. 481–557.
- MELTON, J.; SIMON, A.R. (1997): Understanding the New SQL. A Complete Guide. San Francisco, Morgan Kaufmann.
- MYRACH, T. (2005): Temporale Datenbanken in betrieblichen Informationssystemen. Prinzipien, Konzepte, Umsetzung. Teubner, Wiesbaden.
- OMG (Object Management Group) (2005): Unified Modeling Language (UML), version 2.0. Superstructure. Online verfügbar unter <http://www.omg.org/technology/documents/formal/uml.htm>, zuletzt geprüft am 7.09.2006.
- OMG (Object Management Group) (2006): Data Warehousing, CWM™ and MOF™ Resource Page. Online verfügbar unter <http://www.omg.org/technology/cwm/>, zuletzt geprüft am 28.12.2007.
- PEDERSEN, H. H. MAEGAARD E.; BLOCK, N. (1997): The Danish Integrated Farm Management System. Proceedings of the First European Conference for Information Technology in Agriculture, Copenhagen.
- PENDSE, N. (2005): The OLAP-Report: What is OLAP? Business Application Research Center, Online verfügbar unter <http://www.olapreport.com/fasmi.htm>, zuletzt geprüft am 28.12.2007.
- PICOT, A.; REICHWALD, R.; WIGAND, R.T. (2001): Die grenzenlose Unternehmung. Information, Organisation und Management. 4. Aufl., Gabler, Wiesbaden.
- PIETERSMA, D.; LACROIX, R.; WADE, K. M. (1998): A Framework for the Development of Computerized Management and Control Systems for Use in Dairy Farming. Journal of Dairy Science, Jg. 81, Nr. 11, S. 2962–2972.
- POOLE, J.; CHANG, D.; TOLBERT, D.; MELLOR, D. (2003): Common Warehouse Metamodel. Developers's Guide. Wiley, Indianapolis.
- RATSCHOW, J. P. (2004): Precision Livestock Farming. DLG-Verlag, Frankfurt.

- REENTS, R.; MASCHKA, R.; GABRIEL, H.W. (2003): netRind - ein internetgestütztes Informationssystem für die Rinderzucht. In: Budde, H.-J.; Müller, R.A.E.; Birkner, U. (Hrsg.): Referate der 24. GIL-Jahrestagung in Göttingen 2003 (Band 16), S. 112–115.
- RICHTER, C. (2004): Einführung in die Biometrie. Grundbegriffe und Datenanalyse. Saphir, Ribbesbüttel.
- SAPIA, C.; BLASCHKA, M.; HÖFLING, G.; DINTER, B. (1998): Extending the E/R Model for the multidimensional paradigm. Proceedings of the 1st Intl. Workshop on Data Warehouse and Data Mining (DWDM'98), Lecture Notes in Computer Science, Vol. 1552, S. 105–116.
- SAPIA, C.; BLASCHKA, M.; HÖFLING, G. (1999): An Overview of Multidimensional Data Models for OLAP. Report. FORWISS, Bayrisches Forschungszentrum für wissensbasierte Systeme.
- SCHHELP, J. (1999): Konzeptuelle Modellierung mehrdimensionaler Datenstrukturen. In: Chamoni, P.; Gluchowski, P. (Hrsg.): Analytische Informationssysteme. Data Warehouse, On-Line Analytical Processing, Data Mining. Berlin. Springer, S. 281–305.
- SCHIEFER, G. (2002): Agribusiness und Produktionsketten. In: Doluschitz, R.; Spilke, J. (Hrsg.): Agrarinformatik. UTB. Stuttgart, S. 327–349.
- SCHLÜNSEN, D. (1988): Versuchsergebnisse zur Rechnergestützten Tierüberwachung. In: Bundesforschungsanstalt für Landwirtschaft (Hrsg.): Elektronikeinsatz in der Tierhaltung. Landbauforschung Völkenrode, Sonderheft 93. Braunschweig-Völkenrode, S. 124–139.
- SCHÖN, H. (1993): Rechnergestützte Produktion. In: Schön, H. (Hrsg.): Elektronik und Computer in der Landwirtschaft. Stuttgart. Ulmer, S. 13–25.
- SCHUBERT, M. (2004): Datenbanken. Theorie, Entwurf und Programmierung relationaler Datenbanken. Teubner, Wiesbaden.
- SCHULZE, C.; SPILKE, J.; LEHNER, W. (2004): Informationsbedarfsanalyse als Grundlage der Datenmodellierung im Rahmen des Precision Dairy Farming. In: Schiefer, G.; Wagner, P.; Morgenstern, M.; Rickert, U. (Hrsg.): Referate der 25. GIL-Jahrestagung, Bonn. Integration und Datensicherheit - Anforderungen, Konflikte und Perspektiven, S. 323–326.
- SCHULZE, C.; SPILKE, J.; LEHNER, W. (2005): Hybrid modeling of operational and analytical data at precision dairy farming. Proceedings EFITA/WCCA 2005 in Vila Real, Portugal. Univ. de Trás-os-Montes e Alto Douro. Vila Real, Portugal, S. 960–965.
- SCHULZE, C.; WOLF, S.; SPILKE, J. (2007): Informationstechnische Voraussetzungen für landwirtschaftliche Unternehmensvergleiche - dargestellt an einem Beispiel von Gesundheitsdaten beim Milchrind. In: Böttinger, S.; Theuvsen, L.; Rank, S.; Morgenstern, M. (Hrsg.): Agrarinformatik im Spannungsfeld zwischen Regionalisierung und globalen Wertschöpfungsketten. Referate der 27. GIL-Jahrestagung. LNI, P-101, S. 191–194.

- SCHWAIBERGER, R. (2004): Software Lohnunternehmer. In: Kuratorium für Technik und Bauwesen in der Landwirtschaft e. V. (Hrsg.): Managementsystem für den ortsspezifischen Pflanzenbau. Verbundprojekt pre agro, Abschlussbericht, Kap. 4, KTBL, Darmstadt. S. 71-92.
- SCHWECKE, E.; WALTER, K. (1990): Integriertes Kontroll- und Managementsystem für landwirtschaftliche Betriebe. In: Deutsche Landwirtschafts-Gesellschaft e.V. (Hg.): Integrierte Programme für landwirtschaftliche Betriebe. Eine Zwischenbilanz. Arbeitsunterlagen B/90. DLG-Verlag, Frankfurt. S. 43–60.
- SCHWICKERT, A.C.; FISCHER, K. (1996): Der Geschäftsprozeß als formaler Prozeß - Definition, Eigenschaften, Arten. Arbeitspapiere WI, 4/1996, Lehrstuhl für Allgemeine BWL und Wirtschaftsinformatik, Universität Mainz.
- SEIFERT, B. (1993): Entscheidungsunterstützungsmodule für die Milchviehhaltung. DLG-Verlag, Frankfurt.
- SPIPKE, J.; BÜSCHER, W.; DOLUSCHITZ, R.; FAHR, R.-D.; LEHNER, W. (2003): Precision Dairy Farming - integrativer Ansatz für eine nachhaltige Milcherzeugung. Zeitschrift für Agrarinformatik, S. 19–25.
- SPIPKE, J. (2003): Informationssysteme und ihre Vernetzung beim Rind. In: Fahr, R. D.; Lengerken, G. von (Hrsg.): Milcherzeugung. Grundlagen - Prozesse - Qualitätssicherung. Deutscher Fachverlag, S. 541–551.
- STAHLKNECHT, P.; HASENKAMP, U. (2005): Einführung in die Wirtschaftsinformatik. Springer, Berlin.
- STRAUCH, B.; WINTER, R. (2002): Vorgehensmodell für die Informationsbedarfsanalyse im Data Warehousing. In: Maur, E.; Winter, R. (Hrsg.): Vom Data Warehouse zum Corporate Knowledge Center. Physica. Heidelberg, S. 359–378.
- SWENSSON, C.; SEDERBLAD, B. (1997): Data Warehouse - a new tool in extension service to swedish milkproducers. Proceedings of the First European Conference for Information Technology in Agriculture, Copenhagen.
- TANSEL, A.U.; CLIFFORD, J.; GADIA, S.; JAJODIA, S.; SEGEV, A.; SNODGRASS, R. (Hrsg.) (1993): Temporal databases. Theory, design, and implementation. Benjamin-Cummings, Redwood City.
- TESCHKE, M. (1999): Datenkonsistenz in Data Warehouse Systemen. Arbeitsberichte des Institutes für mathematische Maschinen und Datenverarbeitung, Band 32, Nr. 12, Friedrich-Alexander-Universität Erlangen-Nürnberg.
- TOTOK, A.; JAWORSKI, R. (1998): Modellierung von multidimensionalen Datenstrukturen mit ADAPT. Arbeitsberichte des Instituts für Wirtschaftswissenschaften, Nr. 98/11, Technische Universität zu Braunschweig, Braunschweig.
- VOSSEN, G. (1994): Datenmodelle, Datenbanksprachen und Datenmanagementsysteme. Addison-Wesley, Bonn.
- VOB, S.; GUTENSWAGER, K. (2001): Informationsmanagement. Springer, Berlin.
- WAGNER, P.; KUHLMANN, F. (1991): Concept and implementation of an integrated decision support system (IDSS) for capital-intensive farming. Agricultural Economics, Jg. 5, S. 287–310.

- WENDT, K.; GEORG, D. (2001): Data-Warehouse-Lösungen als Basis der Entscheidungsunterstützung in landwirtschaftlichen Unternehmen. In: Kögl, H.; Spilke, S.; Birkner, U. (Hg.): Referate der 22. GIL-Jahrestagung in Rostock 2001 (Band 14), S. 155–158.
- WENDT, K.; DOLUSCHITZ, R. (2002): Informations- und Kommunikationsmanagement. In: Doluschitz, R.; Spilke, J. (Hrsg.): Agrarinformatik. UTB. Stuttgart, S. 166–208.
- WENDT, K.; SPILKE, J.; THIEDE, M.; PIOTRASCHKE, H. (2004): Outsourcing von IV-Aufgaben landwirtschaftlicher Unternehmen - Einordnung und Nutzungsperspektiven. Zeitschrift für Agrarinformatik, Jg. 12, Nr. 2, S. 34–42.
- WIGGANS, G. R. (1994): Meeting the needs at the national level for genetic evaluation and health monitoring. Journal of Dairy Science, Jg. 77, Nr. 7, S. 1976–1983.
- WINTER, R.; STRAUCH, B. (2004): Information requirements engineering for data warehouse systems. Proceedings of the 2004 ACM symposium on Applied computing, Nicosia, Cyprus. ACM Press. New York, S. 1359–1365.
- ZEHNDER, C. A. (2005): Informationssysteme und Datenbanken. 8. Aufl., vdf Hochschulverlag, Zürich.
- ZHUGE, Y.; GARCIA-MOLINA, H.; WIENER, J. L. (1998): Consistency Algorithms for Multi-Source Warehouse View Maintenance. Distributed and Parallel Databases, Jg. 6, Nr. 1, S. 7–40.

Glossar

Ableitungsvorschrift

Definition zur Berechnung von \rightarrow *Kennzahlen* aus \rightarrow *Fakten* oder abgeleiteten Attributen aus originären Attributen.

ADIS/ADED

ADIS/ADED ist ein Standard zum Datenaustausch in der Landwirtschaft. ADIS (ISO 11787) beschreibt die Form der Datenübertragung mittels ASCII-Daten, während ADED (ISO 11788-1) Datenelemente im Data Dictionary definiert.

Assertion

Assertions (Zusicherungen) stellen einen Mechanismus zur Kontrolle definierter Bedingungen in Datenbanken dar. Sie sind dabei im Gegensatz zu CHECK-Constraints nicht an Tabellen oder Spalten gebunden (MELTON und SIMON, 1997, S. 211 ff.).

CHECK-Constraint

Mechanismus zur Definition von tabellen- oder spaltengebundenen Regeln mithilfe der SQL (MELTON und SIMON, 1997, S. 211 ff.).

Data Dictionary

Metadatenbank eines Datenbankmanagementsystems zur Verwaltung von Datenbanken.

Data Warehouse

Ein Data Warehouse ist eine themenbezogene, integrierte, zeitvariierende, nicht flüchtige Sammlung von Daten zur Unterstützung von Management-Entscheidungsprozessen.

Data Warehouse-Prozess

Der Begriff benennt den Prozess des Betreiben eines \rightarrow *Data Warehouses* mit den Schritten Datenbeschaffung (\rightarrow *ETL*) und Datenspeicherung. Die Analyse der Daten wird in dieser Arbeit nicht als Teil des Data Warehouse-Prozesses gesehen (vgl. dagegen BAUER und GÜNZEL, 2004, S. 526).

Datenbank

Eine Datenbank ist eine strukturierte Sammlung von logisch zusammenhängenden Daten.

Datenbankmanagementsystem

Ein Datenbankmanagementsystem ist ein Softwaresystem zum Erstellen, Pflegen, Abfragen und Manipulieren von \rightarrow *Datenbanken*.

Datenbanksystem

Als Datenbanksystem bezeichnet man →*Datenbankmanagementsysteme* inklusive der von ihnen verwalteten →*Datenbanken*.

Datenorganisation

Unter dem Begriff der Datenorganisation werden alle Verfahren zur Strukturierung (logische Datenorganisation) und Speicherung (physische Datenorganisation) von Daten zusammengefasst (STAHLKNECHT und HASENKAMP, 2005, S. 131).

Dimension

Die Menge der Dimensionen stellt den qualifizierenden Teil eines multidimensionalen Modells dar. Eine Dimension besteht aus einer Menge von Attributen, nach denen →*Fakten* und →*Kennzahlen* angeordnet werden (eindeutige, orthogonale Strukturierung des Datenraumes).

Expertensystem

Softwaresystem zur Informationsgenerierung durch das Anwenden von Regeln auf vorhandene Daten. Die Regeln bilden dabei Expertenwissen der entsprechenden Domäne ab (PICOT et al., 2001, S. 171 f.). Expertensysteme gelten als Systeme der künstlichen Intelligenz.

ETL (Extract-Transform-Load)

Datenaufarbeitungsprozess im Rahmen des →*Data Warehouse-Prozesses* mit den Schritten der Extraktion von Daten aus Quellsystemen, Transformation der Daten in die Zielstrukturen und das Laden ins Data Warehouse.

Fakt

Fakten stellen quantifizierende Daten dar, die in einer bestimmten Granularität vorliegen. (Sie entsprechen der Ausprägung quantifizierender Attribute im E/R-Modell.)

Factless Fact

Eine Factless Fact-Tabelle weist die Struktur einer Faktentabelle im multidimensionalen Kontext, jedoch ohne Kennzahl-Attribut auf.

Funktionale Abhängigkeit

Eine Attributmenge A des Relationsschemas R ist funktional abhängig von einer Attributmenge B des gleichen Relationsschemas R, wenn zu jeder Wertekombination von B genau eine Wertekombination von A existiert.

Granularität

Die Granularität bezeichnet den Verdichtungsgrad von Daten. Je niedriger die Granularität, desto detaillierter sind die Daten.

Informationssystem

Ein Informationssystem ist ein System, das große Mengen von Daten strukturiert, effizient und dauerhaft speichert, Anfragen und Änderungen ausführt (BISKUP, 1995, S. 2).

Integrität

Datenintegrität umfasst die Bereiche Datenkonsistenz (\rightarrow *Konsistenz*), Datensicherheit (Sicherung gegen Verlust und Verfälschung) und Datenschutz (Schutz vor unbefugtem Zugriff) (ZEHNDER, 2005, S. 244; LASSMANN et al., 2001, S. 232).

Integritätsbedingungen

Bedingungen, die die Integrität von Daten definieren und deshalb beim Datenbankentwurf festzulegen sind.

Kardinalität

Kardinalitäten beschreiben die Komplexität von Beziehungstypen zwischen Entitätstypen, das heißt sie geben an, mit wie vielen Entitäten eines anderen Entitätstyps eine Entität in Beziehung stehen muss oder kann.

Kennzahlen (=abgeleitete Kennzahlen, =measure):

Kennzahlen sind \rightarrow *quantifizierende Daten*, die nach einer \rightarrow *Ableitungsvorschrift* aus \rightarrow *Fakten* berechnet werden. (z.B. Anzahl Mastitisbehandlungen = Anzahl Behandlungen mit Diagnose ‚Mastitis‘). Ein Spezialfall ist Kennzahl = Fakt, z.B. bei der Milchmenge.

Konsistenz

Konsistenz ist die Widerspruchsfreiheit von Daten (ZEHNDER, 2005, S. 243).

Klassifikationshierarchie

Eine Klassifikationshierarchie ist die Ausprägung eines \rightarrow *Klassifikationsschemas*, das heißt die Hierarchie der Instanzen von \rightarrow *Klassifikationsstufen*.

Klassifikationsschema

Ein Klassifikationsschema stellt eine hierarchische Anordnung von Attributen einer \rightarrow *Dimension* dar. Die Attribute werden damit zu \rightarrow *Klassifikationsstufen* (=Hierarchiestufen, =Hierarchieebenen). Sie bildet einen Auswertungspfad (=Navigationpfad). Parallelklassifikationen sind möglich. Weiterhin ist die Einführung von dimensionalen Attributen möglich, die funktional von einem Klassifikationsattribut abhängen.

Klassifikationsstufen

Klassifikationsstufen sind Attribute einer Dimension im Rahmen von \rightarrow *Klassifikationsschemata*.

Materialisierte Sicht

Als materialisierte Sicht wird an dieser Arbeit allgemein die physische Speicherung von Daten verstanden, die prinzipiell auch durch SQL-Views repräsentiert werden könnten.

Metrik

\rightarrow Ableitungsvorschrift

Normalisierung

Die Normalisierung ist ein Prozess zur Eliminierung von Datenabhängigkeiten mittels Zerlegung von Relationen mit dem Ziel der Redundanzvermeidung (VOSSEN, 1994, S. 51; KEMPER und EICKLER, 2004, S. 171).

NULL-Wert

Der Wert NULL drückt einen unbestimmten Zustand eines Attributes aus.

Qualifizierende Daten

Qualifizierende Daten sind beschreibende Daten, die dem Benutzer als Navigationsraster zur Verfügung stehen. Diese können Begriffshierarchien im Sinne von →*Klassifikationsschemata* bilden, die die Bedeutung von Auswertungspfaden annehmen.

Quantifizierende Daten

Quantifizierende Daten sind Daten, die den Auswertungsgegenstand (→*Kennzahlen*) darstellen (z.B. Milchmenge und Anzahl Mastitisbehandlungen).

Redundanz

Redundanz ist das mehrfache Vorhandensein gleicher Daten. Von Redundanz spricht man bei mehrfach vorhandenen, identischen Daten, aber auch bei Daten, die von vorhandenen Daten abgeleitet werden (redundante Information).

Referenzmodell

Ein Referenzmodell ist ein Modell mit dem Anspruch einer gewissen Allgemeingültigkeit für einen bestimmten Anwendungsbereich (LASSMANN et al., 2001, S. 372).

Transaktion

Eine Transaktion stellt eine unteilbare Folge von (Datenmanipulations-)Anweisungen dar.

Trigger

Trigger sind durch Datenmodifikationsoperationen ausgelöste Aufrufe gespeicherter Funktionen.

Anhang A: Referenzmodell Milcherzeugung

Das Referenzmodell Milcherzeugung wurde mit dem Sybase PowerDesigner[®], Version 10 erstellt. Die Darstellung des relationalen Datenschemas (Abbildung A 2 bis Abbildung A 10) erfolgt mithilfe der in Abbildung A 1 abgebildeten Grafikelemente. Die Beschriftung der Relationen und verwendete Datentypen sind in Tabelle A 1 und Tabelle A 2 angeführt.

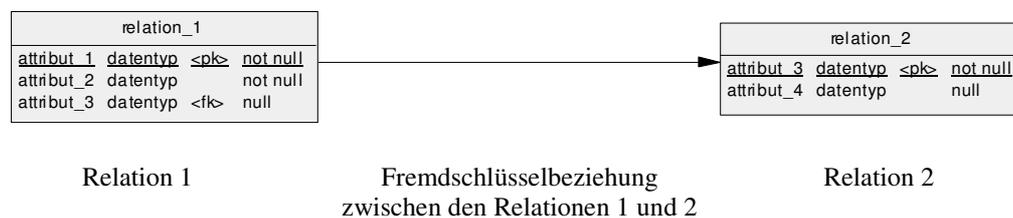


Abbildung A 1: Darstellungselemente des Referenzmodells

Tabelle A 1: Beschriftung im Referenzmodell

Element	Beschreibung
<i>relation_1</i>	Bezeichnung der Relation
<i>attribut_1</i>	Bezeichnung des Attributes
<i>datentyp</i>	Datentyp des Attributes
<i><pk></i>	Kennzeichnung als Primärschlüsselattribut
<i><fk></i>	Kennzeichnung als Fremdschlüsselattribut
<i>not null</i>	nicht-NULL-Bedingung
<i>null</i>	NULL-Wert zulässig

Tabelle A 2: Datentypen des Referenzmodell

Datentyp	Beschreibung
bool	Boole'scher Wert
char(n)	Zeichenkette mit fester Länge n
decimal(m,n)	Dezimalzahl mit m Gesamtstellen und n Nachkommastellen
int4	4-Byte Ganzzahl
serial	selbstzählende 4-Byte Ganzzahl
text	Zeichenkette mit variabler Länge ohne Höchstgrenze
time	Tageszeit
timestamp with time zone	Datum und Zeit mit Zeitzone
varchar(n)	Zeichenkette mit variabler Länge, Maximallänge n

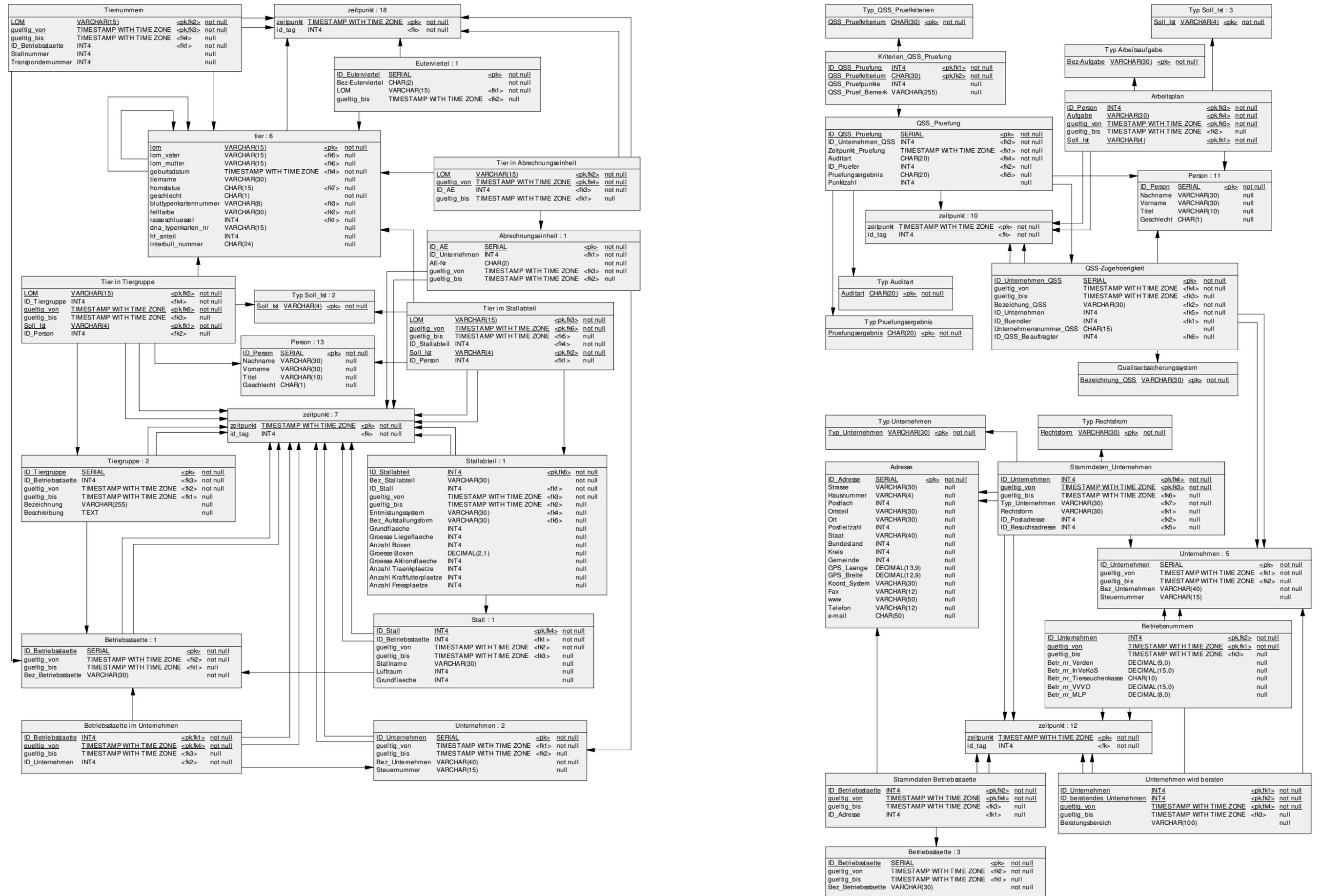


Abbildung A 2: Referenzmodell, Teil Organisationseinheiten

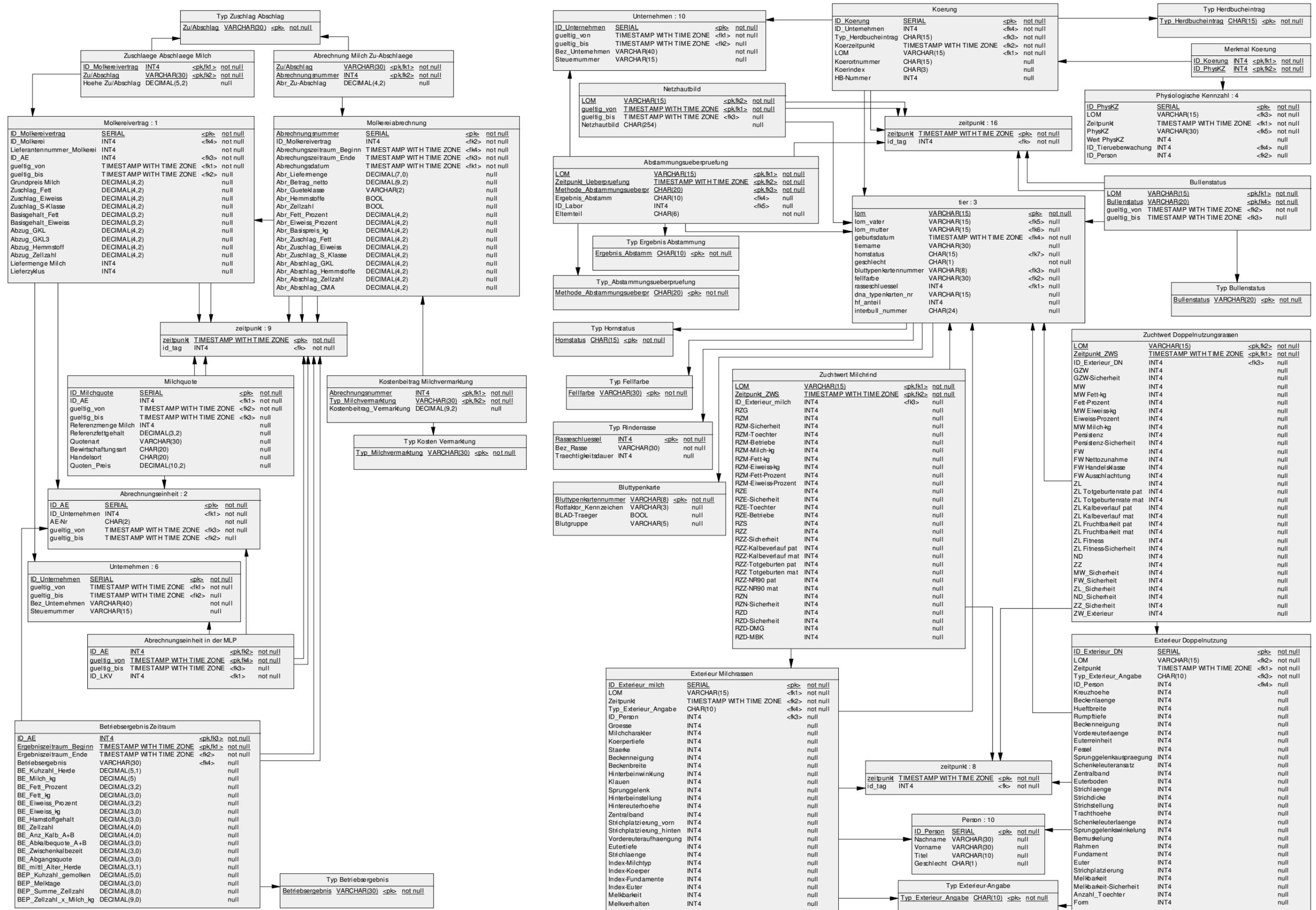


Abbildung A 3: Referenzmodell, Teil Stammdaten Tier und Stammdaten Abrechnungseinheit

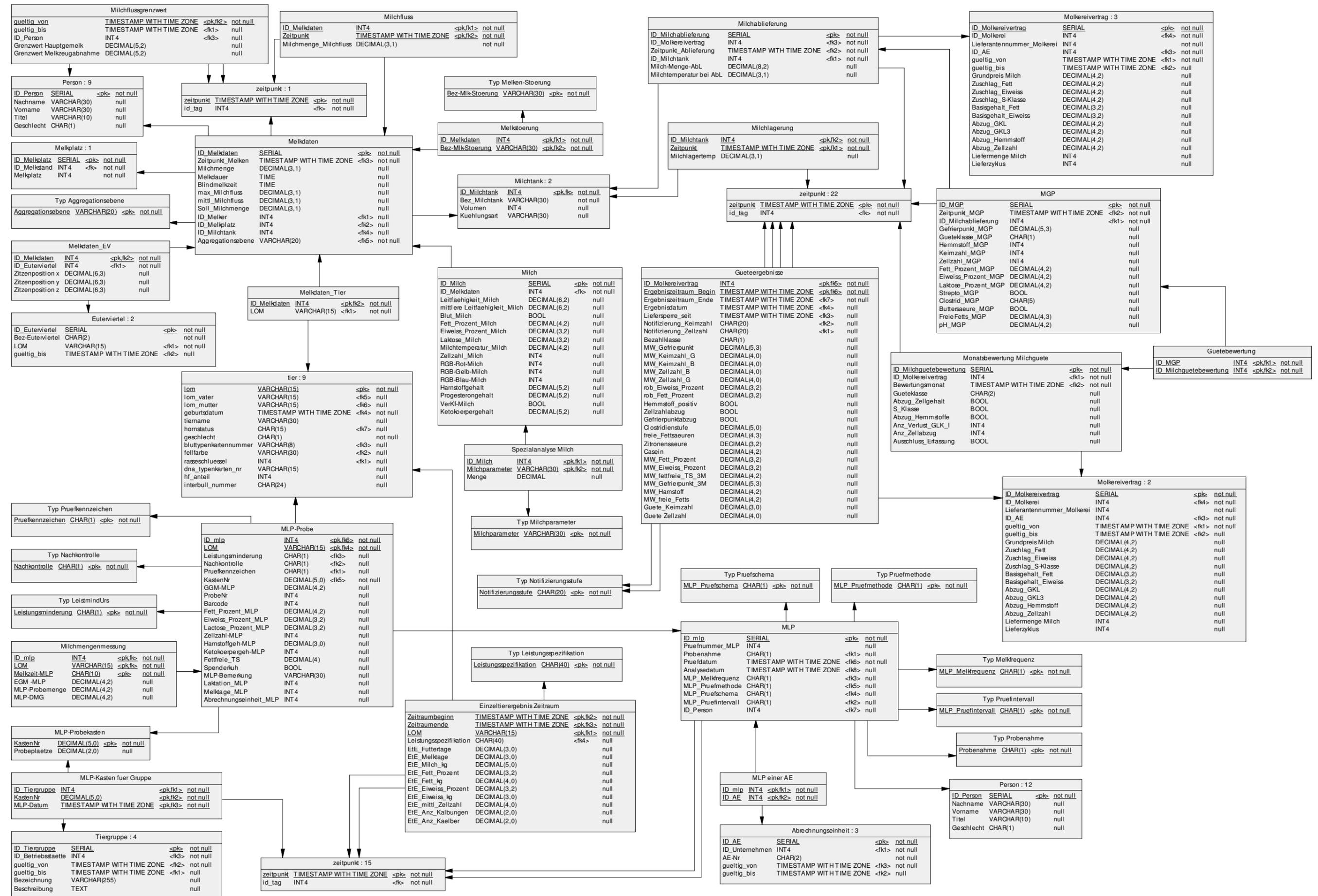


Abbildung A 7: Referenzmodell, Teil Milchgewinnung

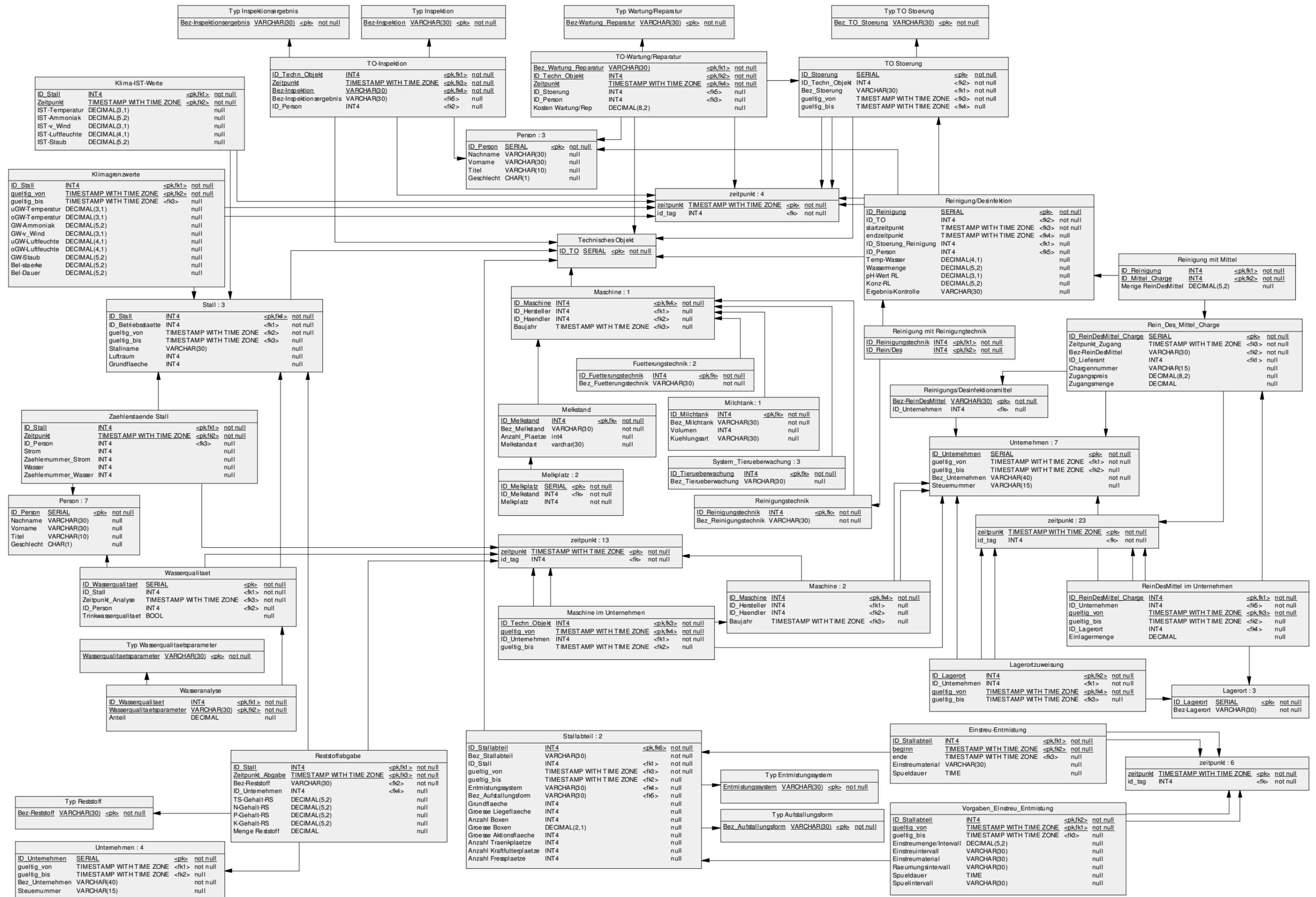


Abbildung A 8: Referenzmodell, Teil Technik

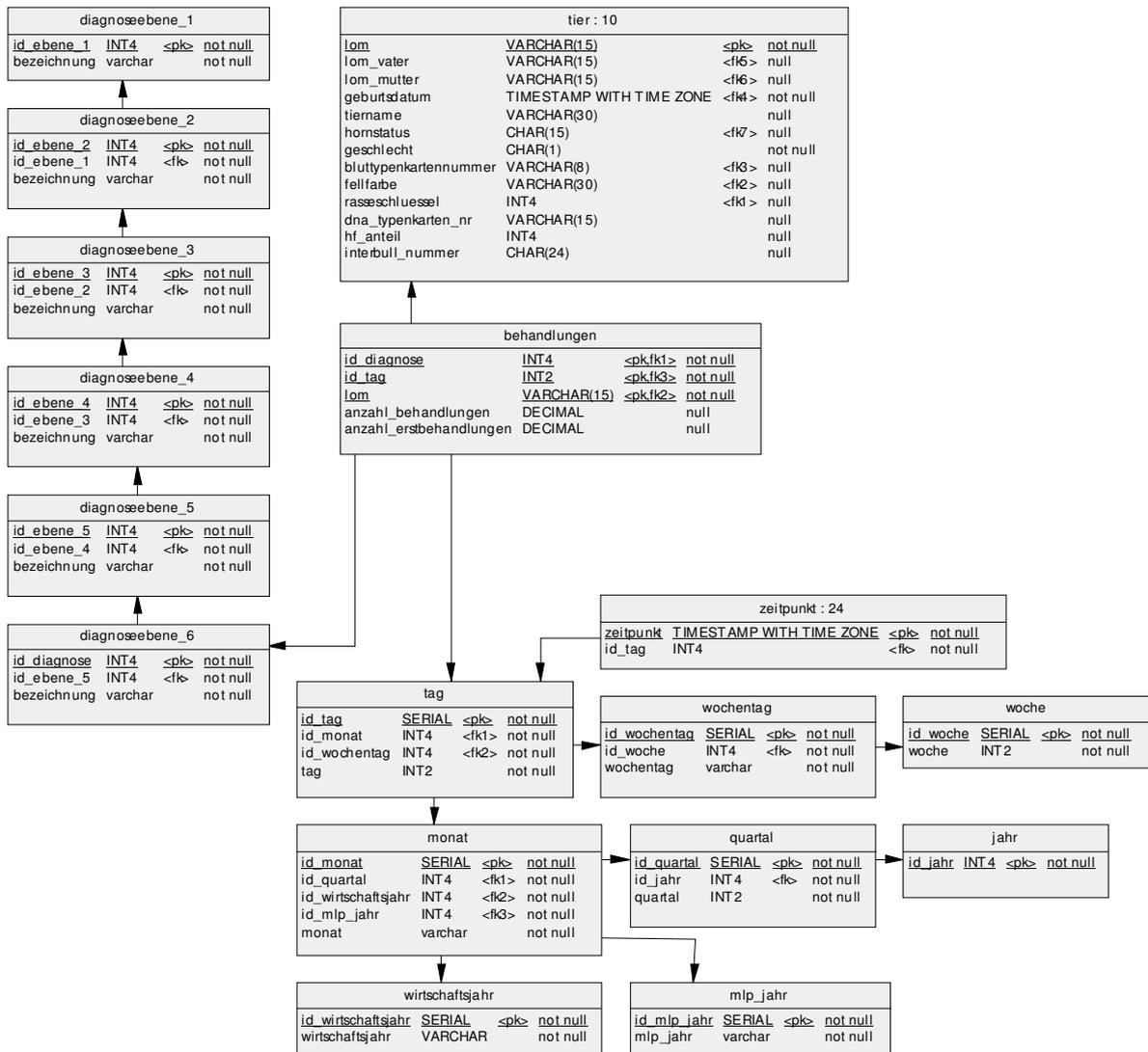


Abbildung A 9: Referenzmodell, Krankheitsbehandlungen, analytisch motivierte Strukturen

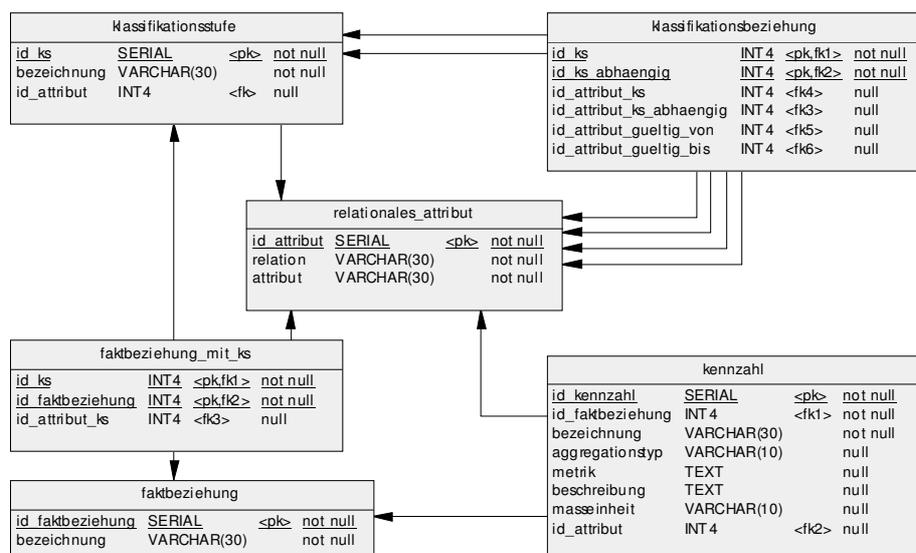


Abbildung A 10: Referenzmodell, Teil Metamodell

Anhang B: Material der Informationsbedarfsanalyse

Tabelle A 3: Teilverfahren, Prozesse und Merkmalskomplexe der Informationsbedarfsanalyse

Teilverfahren	Prozess	Merkmalskomplex
Milchgewinnung	Melkung	Milchqualitätserfassung
		Eutergesundheitserfassung
		Milchabgabe (Mengen und Zeiten)
		Prozessablauf
		Eutergeometrie
	MLP	Probekendaten
		Beprobungsergebnisse
		Tierdaten
	Sekundärdatenerfassung im Melkstand	Aktivitätskennzahlen
		Physiologische Parameter
	Milchlagerung	
	Milchablieferung	
	Dokumentation Melk- und Kühltechnik	Dokumentation technische Ausstattung
Dokumentation Reinigung		
Dokumentation Inspektion und Wartung Melk- und Kühltechnik		
Dokumentation technische Störung		
Fütterung	Futtermittel	Stammdaten
		Futtermittelbeprobung
		Nährwertparameter
		Mineralstoffe
		Vitamine
		Zusätzliche Grundfutterparameter
	Sensorik	
	Rationsberechnung (Soll-Ration)	
	Rationsmischung (Ist-Ration)	
	Futteraufnahme	Kraftfutter
		Grundfutter
Wasseraufnahme		
Erstversorgung Kalb		
Weidegang		
Behandlungen, Gesundheits- und Konditionskontrolle und -erfassung	Tierseuchenstatus	
	Untersuchungen	
	Maßnahmen/Behandlungen	
	Medikamenteinsatz	

Fortsetzung Tabelle A 3: Teilverfahren, Prozesse und Merkmalskomplexe der Informationsbedarfsanalyse

Teilverfahren	Prozess	Merkmalskomplex
Behandlungen, Gesundheits- und Konditionskontrolle und -erfassung	Konditionserfassung	Lebendmassebestimmung
		Body Condition Score (BCS)
		Rückenfett Dicke
	Gesundheitsüberwachung	Tierverhalten
	Aktivitätskennzahlen	
	Physiologische Parameter	
Reproduktion	Zuchtwertschätzung	Milchleistung
		Exterieur
		Zellzahl
		Nutzungsdauer
		Zuchtleistung
	Brunsterkennung	Aktivitätskennzahlen
		Physiologische Parameter
	Brunstinduktion	
	Besamung	
	Embryo Transfer	
Trächtigkeitsuntersuchung		
Kalbung		
Dokumentation, Kennzeichnung, Bestandesorganisation	Tierstammdaten	Tierkennzeichnung
		Geburtsdaten
		Abstammung
		Zugang
		Abgang
	Leistungsdaten	
Tierbewegung		
Einstreu und Entmistung	Einstreu	
	Entmistung	Gülle
		Dung
Klimaerfassung und -steuerung	IST-Klima	
	SOLL-Klima (Grenzwerte)	
	Licht	
Stalltechnik und -gestaltung Produktionsablauf	Bauliche und technische Ausstattung	
	Wartung und Kontrolle technische Einrichtungen	
	Wasserqualitätsüberprüfung	
	Maßnahmen / Behandlungen an Objekten	
	Reinigung / Desinfektion von Objekten	
	Verbrauchsdaten (Energie und Wasser)	
Betriebsstammdaten		

Tabelle A 4: Merkmale und Merkmalsgruppen der Informationsbedarfsanalyse

Objekt	Merkmalsgruppe	Merkmale	
Euterviertel	Melkdaten Euterviertel	LOM_Tier	
		Identifikation_Euterviertel	
		Beginn_Melken	
		Ende_Melken	
		Temperatur_Milch	
		Gemelksmenge	
		Soll_Gemelk	
		Leitfähigkeit_Milch	
		Fett%	
		Eiweiß%	
		Laktose%	
		Zellzahl	
		Zitzenposition	
		Farbwert_Milch	
Einzeltier	Melkdaten Tier	LOM_Tier	
		Melkplatz	
		Beginn_Melken	
		Milch_verkehrsfähig	
		Störung_Melken	
		Progesterongehalt_Milch	
		Harnstoffgehalt_Milch	
		Melker	
		Daten Milchleistungsprüfung (MLP)	Probenahmezeitpunkt_MLP
			LOM_Tier
	Probenummer		
	Kastenummer		
	Barcode_Probenflasche		
	Prüfkennzeichen		
	Nachkontrolle		
	Melkzeit		
	Prüfmethode		
	Prüfschema		
	Prüfintervall		
	Melkfrequenz		
	Leistungsmindernde_Ursache		
	Einzelgemelksmenge		
	Gesamtgemelksmenge		
	Fett%		
	Eiweiß%		
	Laktose%		
	Zellzahl		
Harnstoff			
Ketokörpergehalt			

Fortsetzung Tabelle A 4: Merkmale und Merkmalsgruppen der Informationsbedarfsanalyse

Objekt	Merkmalsgruppe	Merkmale
Einzeltier	Aktivitätsdaten	Zeitpunkt_Datenerfassung
		LOM_Tier
		Bewegungskennzahl
		Liegekennzahl
		Erfassungssystem
	Physiologische Kennzahlen	Zeitpunkt_Datenerfassung
		LOM_Tier
		Kennzahlwert
		Kennzahltyp
		Erfassungssystem
		Erfassende_Person
	Daten Kraftfutterzuteilung	LOM_Tier
		SOLL_Menge_Kraftfutter
	Daten Kraftfutterraufnahme	LOM_Tier
		IST-Menge_Kraftfutter
	Brunst	Zeitpunkt_Brunst
		LOM_Tier
		Brunstausprägung
		Erfassende_Person
		Erfassungssystem
	Brunstinduktion	Zeitpunkt_Brunstinduktion
		LOM_Tier
		Hormon
		Hormonmenge
		Verabreichende_Person
	Besamung	Zeitpunkt_Besamung
		LOM_Tier
		Belegungsart
		Identifikation_Bulle
		Identifikation_Spermaportion
		Durchführende_Person
	Embryotransfer (ET)	Zeitpunkt_ET
LOM_Empfängertier		
LOM_Spendertier		
Identifikation_Bulle		
Zeitpunkt_Spülung		
Durchführende_Person		
Trächtigkeits- untersuchung (TU)	Zeitpunkt_TU	
	LOM_Tier	
	TU-Ergebnis	
	Durchführende_Person	

Fortsetzung Tabelle A 4: Merkmale und Merkmalsgruppen der Informationsbedarfsanalyse

Objekt	Merkmalsgruppe	Merkmale
Einzeltier	Kalbung	Beginn_Kalbung
		Ende_Kalbung
		LOM_Tier
		Anzahl_Kälber
		Gesundheitsstatus_Kalb
		Geburtsmasse_Kalb
		Kalbeverlauf
		Art_der_Geburtshilfe
		Person_Geburtshilfe
		Stammdaten
	Tiername	
	Transponder_Nummer	
	Stallnummer	
	Herdbuchnummer	
	Strichzahl	
	Zugehörige_Abrechungseinheit	
	Geburtsdatum	
	Geschlecht	
	Rasse	
	LOM_Mutter	
	LOM_Vater	
	Herkunftsbetrieb	
	Zeitpunkt_Zugang	
	Zeitpunkt_Abgang	
	Abnehmer	
	Abgangsgrund	
	Abgangsursache	
	Person_Abgang_verfügt	
	Schlachtparameter	
		Ausschlachtung
		Handelsklasse
	Tierbewegung	LOM_Tier
		Zugehörige_Tiergruppe
		Zeitpunkt_Bewegung
		Person
	Zuchtwertschätzung	LOM_Tier
		Zeitpunkt_Zuchtwertschätzung
		RZG (Relativzuchtwert Gesamt)
		RZM (Relativzuchtwert Milch)
		RZM_Sicherheit
RZE (Relativzuchtwert Exterieur)		
RZE_Sicherheit		
RZS (Relativzuchtwert somatischer Zellgehalt)		
RZS_Sicherheit		

Fortsetzung Tabelle A 4: Merkmale und Merkmalsgruppen der Informationsbedarfsanalyse

Objekt	Merkmalsgruppe	Merkmale
Einzeltier	Zuchtwert-schätzung	RZN (Relativzuchtwert Nutzungsdauer) RZN_Sicherheit RZZ (Relativzuchtwert Zuchtleistung) RZZ_Sicherheit RZZ_Kalbeverlauf RZZ_Totgeburten RZZ_NR90 RZD (Relativzuchtwert Melkbarkeit) RZD_Sicherheit RZD_durchschnittliches_Minutengemelk RZD_Melkbarkeit RZM_Milch_kg RZM_Fett_kg RZM_Eiweiß_kg RZM_Fett% RZM_Eiweiß% Exterieurmerkmal_Milchcharakter Exterieurmerkmal_Größe Exterieurmerkmal_Körpertiefe Exterieurmerkmal_Stärke Exterieurmerkmal_Beckenneigung Exterieurmerkmal_Beckenbreite Exterieurmerkmal_Hinterbeinwinkelung Exterieurmerkmal_Klauen Exterieurmerkmal_Sprungelenk Exterieurmerkmal_Hinterbeinstellung Exterieurmerkmal_Hintereuterhöhe Exterieurmerkmal_Zentralband Exterieurmerkmal_Strichplazierung_vorn Exterieurmerkmal_Strichplazierung_hinten Exterieurmerkmal_Vordereuteraufhängung Exterieurmerkmal_Eutertiefe Exterieurmerkmal_Strichlänge Index_Milchtyp Index_Körper Index_Fundamente Index_Euter Einzelmerkmal_Melkbarkeit Einzelmerkmal_Melkverhalten Zeitpunkt_Bonitur_Exterieur Bonitierende_Person
Tiergruppe	Stammdaten	Identifikation_Tiergruppe Identifikation_zugehöriger_Stall Zeitpunkt_Zuordnung

Fortsetzung Tabelle A 4: Merkmale und Merkmalsgruppen der Informationsbedarfsanalyse

Objekt	Merkmalsgruppe	Merkmale
Tiergruppe	Futtrationzuweisung	Zeitpunkt_Gültigkeit
		Soll_Menge_Ration
		Identifikation_Ration
	Futtrationvorlage	Vorlagemenge
		Identifikation_Ration
		Zeitpunkt_Rationsmischung
		Restmenge_Futtrvorlage
		Zeitpunkt-Vorlage
		Vorlegende_Person
		Vorlegende_Maschine
	Weidedaten	Zeitpunkt-Auftrieb
		Zeitpunkt-Abtrieb
		Identifikation_Weidefläche
	Einstreudaten	Einstreumaterial
		Einstreuintervall
		Einstreumenge_je_Intervall
	Entmistung	Soll_Intervall_Entmistung/Spülung
		Beginn_Entmistung/Spülung
		Ende_Entmistung/Spülung
Stall/Stallabteil	Reststoffabgabe	Zeitpunkt_Reststoffabgabe
		Reststoffart
		Reststoffmenge
		Abnehmer
		TS-Gehalt
		N-Gehalt
		P-Gehalt
		K-Gehalt
	Klimadaten	Zeitpunkt_Klimamessung
		Ist_Temperatur
		Ist_NH3
		Ist_Windgeschwindigkeit
		Ist_Luftfeuchte
		Ist_Staub
	Klimagrenzwerte (GW)	Gültigkeitszeitpunkt_Klimagrenzwerte
		Grenzwert_Temperatur
		Grenzwert_Ammoniak
		Grenzwert_Windgeschwindigkeit
		Grenzwert_Luftfeuchte
		Grenzwert_Staub
		Beleuchtungsstärke
		Beleuchtungsdauer
		Verfügende_Person

Fortsetzung Tabelle A 4: Merkmale und Merkmalsgruppen der Informationsbedarfsanalyse

Objekt	Merkmalsgruppe	Merkmale	
Stall/Stallabteil	Technische Ausstattung	Gültigkeitszeitpunkt	
		Anzahl_Boxen	
		Größe_Boxen	
		Aufstellungsform	
		Entmistungssystem	
		Anzahl_Abkalbeboxen	
		Größe_Abkalbebox	
		Anzahl_Fressplätze	
		Anzahl_Kraftfutterplätze	
		Anzahl_Tränken	
		Luftraum	
		Grundfläche	
		Größe_Aktionsfläche	
	Wasserqualität		Zeitpunkt_Untersuchung
			Trinkwasser_Qualität
			Qualitätsmangel
Verbrauchsdaten Stall		Zeitpunkt_Ablesen_Strom	
		Zählerstand_Strom	
		Zeitpunkt_Ablesen_Wasser	
		Zählerstand_Wasser	
Stammdaten		Identifikation_Stall	
		Identifikation_zugeordnetes_Unternehmen	
		Gültigkeitszeitpunkt_Zuordnung	
Abrechnungseinheit (AE)	Stammdaten	Identifikation_Abrechnungseinheit	
		Gültigkeitszeitpunkt	
		Identifikation_zugeordnetes_Unternehmen	
		Milchquotenmenge	
		Referenzfettgehalt_Milchquote	
	Daten Milchablieferung / Milchgüte		Zeitpunkt_Milchablieferung
			Abnehmende_Molkerei
			Menge_Ablieferungsmilch
			Milchtemperatur
			Güteklasse
			Keimzahl
			Zellzahl
			Hemmstoffe
	Gefrierpunkt		
	Daten Milchlagerung		Milchtemperatur
Zeitpunkt_Milchtemperaturerfassung			
Betriebszweig/-stätte / Unternehmen	Stammdaten	Identifikation_Betriebsstätte	
		Straße/Hausnummer	
		Ort	
		Postleitzahl	
		Identifikation_Unternehmen	

Fortsetzung Tabelle A 4: Merkmale und Merkmalsgruppen der Informationsbedarfsanalyse

Objekt	Merkmalsgruppe	Merkmale
Betriebszweig/- stätte / Unternehmen	Stammdaten	MLP_Nummer_LKV
		InVeKoS_Unternehmensnummer
		VVVO_Nummer
		Schlüssel_Verden
		Nummer_Tierseuchenkasse
Allgemein	Daten Wartung, Reparatur	Zeitpunkt_Wartung/Reparatur
		Art_der_Wartung/Reparatur
		Reparierende_Person
	Störungsdaten	Beginn_Störung
		Ende_Störung
		Art_der_Störung
		dokumentierende_Person
	Inspektionsdaten	Zeitpunkt_Inspektion
		Art_der_Inspektion
		Ergebnis_Inspektion
		Inspizierende_Person
	Stammdaten Maschine	Identifikation_Maschine
		Baujahr
	Daten Reinigung, Desinfektion	Reinigungsgegenstand
		Angewandtes_Mittel
		Wassertemperatur
		Wassermenge
		pH-Wert_Lösung
		Konzentration_Lösung
		Beginn_Reinigung
		Ende_Reinigung
		Ergebnis_Kontrolle
		Reinigende_Person
		Verwendete_Technik
		Störung_Reinigung
	Lager	Zeitpunkt_Lagerbestand
		Bezeichnung_Lagerartikel
		Chargennummer
		Ist_Menge
		Person_Inventur
	Futtermittel (FM)	Identifikation_Futtermittel
		Futtermittelart
		Zugangsdatum
Zugangsmenge		
Hersteller		
Lieferant		
Chargennummer		
Zeitpunkt_Probenahme		
Probennummer		

Fortsetzung Tabelle A 4: Merkmale und Merkmalsgruppen der Informationsbedarfsanalyse

Objekt	Merkmalsgruppe	Merkmale
Allgemein	Futtermittel	Probenart (Frisch/Konservat)
		Person_Probenahme
		Zeitpunkt_Futtermittelanalyse
		Analysierendes_Labor
		Trockensubstanzgehalt
		Rohasche
		Organische_Masse
		Verdauliche_organische_Masse
		Rohprotein
		Verdauliches_Rohprotein
		Unabbaubares_Rohprotein
		Proteinlöslichkeit
		Durchflussprotein
		Nutzbares_Rohprotein
		Rohfett
		Verdauliches_Rohfett
		Rohfaser
		Strukturwert
		Strukturwirksame_Rohfaser
		Verdauliche_Rohfaser
		N-freie_Extraktstoffe (NfE)
		Verdaulichkeit_NfE
		NDF (neutral detergent fiber)
		ENDF (effective neutral detergent fiber)
		ADF (acid detergent fiber)
		ADL (acid detergent lignin)
		Stärke
		Beständige_Stärke
		Zucker
		RNB (Ruminale N-Bilanz)
		NEL (Nettoenergie-Laktation)
		Ca-Gehalt (Kalzium)
		P-Gehalt (Phosphor)
		Na-Gehalt (Natrium)
		Mg-Gehalt (Magnesium)
		K-Gehalt (Kalium)
		Cl-Gehalt (Chlor)
		S-Gehalt (Schefel)
		Vitamin A-Gehalt
		Carotin-Gehalt
Vitamin E		
Toxinart		
Toxinmenge		
Werbungsart (Schlag)		

Fortsetzung Tabelle A 4: Merkmale und Merkmalsgruppen der Informationsbedarfsanalyse

Objekt	Merkmalsgruppe	Merkmale
Allgemein	Futtermittel	Zeitpunkt_Ernte
		Siliermittel
		Menge_Siliermittel
		Lagerort
		Zeitpunkt_Silobeginn
		Schnittkennzeichen
		Schnittzeitpunkt
		Farbe
		Geruch
		Fremdbesatz
		Schimmelbonitur
		pH_Wert
		Ammoniak
		Buttersäure
		Essigsäure
	Rationsberechnung	Zeitpunkt_Rationsberechnung
		Identifikation_Ration
		Identifikation_Einzelfuttermittel
		Menge_Einzelfuttermittel
		Berechnende_Person
	Rationsmischung	Zeitpunkt_Rationsmischung
		Identifikation_Ration
		Identifikation_Einzelfuttermittel
		Menge_Einzelfuttermittel
		Mischende_Person
	Medikament	Bezeichnung_Medikament
		Indikation
		Verabreichungsart
		Dosis
		Einheit
		Wartezeit_Milch
		Wartezeit_Impfen
		Warezeit_Fleisch
		Nummer_Arzneimittelbeleg
		Zeitpunkt_Arzneimittelabgabe
		Medikament_Charge
	Behandlungen	Zeitpunkt_Behandlung
		Behandlungsgegenstand
		Behandlungsgrund
		Behandlungsart
		Behandlungsmittel
Menge_Arzneimittel		
Verabreichungsart		
Behandlungserfolg		

Fortsetzung Tabelle A 4: Merkmale und Merkmalsgruppen der Informationsbedarfsanalyse

Objekt	Merkmalsgruppe	Merkmale
Allgemein	Behandlungen	Diagnose
		Behandelnde_Person
	Untersuchung	Zeitpunkt_Untersuchung
		Untersuchungsgegenstand
		Untersuchungsgrund
		Zeitpunkt_Voruntersuchung
		Untersuchungsart
		Befund/Diagnose
		Zeitpunkt_Befund
		Untersuchende_Person

Danksagung

Diese Arbeit entstand im Rahmen eines durch die Deutsche Forschungsgemeinschaft (DFG) geförderten Projektes.

Darüber hinaus gab es zahlreiche Personen, die zum Gelingen dieser Arbeit beigetragen haben:

Prof. Dr. J. Spilke und Prof. Dr. W. Lehner sei gedankt für die Überlassung des Themas, für Hinweise und Anregungen.

Weiterhin danke ich allen ehemaligen Kollegen der Arbeitsgruppe Biometrie und Agrarinformatik. Stellvertretend möchte ich Peggy Walther nennen, die unermüdlich auftretende Fragen beantwortete.

Zu danken habe ich weiterhin Andreas und Michael Schulze, Heike Böttcher und Kerstin Ziehe für die Durchsicht der vorliegenden Arbeit oder Teilen davon.

Ausdrücklicher Dank gebührt meinen Eltern Erhard und Gisela Schulze, grundsätzlich für die gewährte Freiheit bei meiner Ausbildung, insbesondere jedoch für die berufliche Entlastung in der letzten Phase der Arbeit.

Danke.

Wissenschaftliche Entwicklung des Bewerbers

- 10/1997 – 04/2003 Studium der Agrarwissenschaften an der Humboldt-Universität zu Berlin, Abschluss: Diplom-Agraringenieur
Arbeitsgebiet der Diplomarbeit: Zerkleinerung von Futtergetreide
- 12/1999 - 10/2002 Praktika und mehrmalige Beschäftigung als studentische Hilfskraft am Institut für Agrartechnik Bornim e.V., Potsdam:
- Abteilung Technik in der Pflanzenproduktion,
Arbeitsgebiet: Teilflächenspezifischer Pflanzenbau
 - Abteilung Aufbereitung, Lagerung und Konservierung,
Arbeitsgebiet: Zerkleinerung von Futtergetreide
- 12/2003 – 11/2006 Wissenschaftlicher Mitarbeiter (50%) an der Martin-Luther-Universität Halle, Landwirtschaftliche Fakultät, Arbeitsgruppe Biometrie und Agrarinformatik
- Bearbeitung des DFG-Projektes „Erstellung und Evaluierung eines Referenzmodells zur Organisation operativer und analytischer Daten als Basis für ein effizientes betriebliches Informationssystem“

Erklärung

Hiermit erkläre ich, Christian Schulze, diese wissenschaftliche Arbeit selbständig angefertigt und keine anderen als die angegebenen Hilfsmittel benutzt zu haben.

Welsickendorf, 21.12.2007