

# Wie erziehe ich meinen Kunden

**Dr. Andrea Kennel**  
**InfoPunkt Kennel GmbH**  
**CH-8600 Dübendorf**

## **Schlüsselworte:**

Data Warehousing, Dimensionen, Fakten, Zusammenarbeit, Kommunikation.

## **Ausgangslage**

Die Informatikabteilung entwickelt und wartet die Anwendungen der Informatik, die vom Business eingesetzt werden. Daher kennt die Informatik die Datenstrukturen und damit die Daten oft besser als das Business. Das Business kennt dafür die Abläufe der Geschäftsprozesse. Das Business hat klar das Bedürfnis, die Daten aus der Datenbank flexibel abrufen zu können. Schlussendlich werden die Daten ja nicht nur für den operativen Betrieb, sondern auch für Auswertungen gesammelt und benötigt.

So entstehen über die Jahre eine Vielzahl von unterschiedlichen Reports und Datenextrakten, die vom Business für eine bessere Übersicht beispielsweise in Excel geladen und weiter verarbeitet werden.

Mit der Zeit entstehen viele punktuelle Insellösungen, die nicht wirklich zusammen passen. Einerseits, weil jede Lösung anders aufgebaut ist und schlimmer, weil jede Lösung andere Zahlen liefert.

Meist kann sich das Business nicht vorstellen, was ein Data Warehouse ist und leisten kann. Falls sie schon mal etwas von einem Data Warehouse gehört haben, dann entweder von einem guten Verkäufer, der ein Data Warehouse als Lösung aller Probleme anpreist. Oder sie haben von einem Kollegen gehört, dass ein Data Warehouse ein Informatikprojekt ist, das viel kostet und selten fertig wird.

Das Ziel ist natürlich, die Anwender von den Vorteilen eines Data Warehouses zu überzeugen, doch muss deutlich werden, dass ein Data Warehouse nie wirklich alle Probleme lösen kann, zumal die meisten gar nicht bekannt sind.

Damit ein Data Warehouse Projekt gelingt, ist eine gute Zusammenarbeit zwischen der Informatik und dem Business nötig. Gehen wir davon aus, dass das Grundwissen über Data Warehousing in der Informatik vorhanden ist, so geht es darum, wie die Informatik das Business erziehen kann. Dieser Artikel beschreibt, wie wir dabei normalerweise vorgehen.

## **Erziehungsschritt 1**

Was ist ein Cube, das kann man damit?

Das Business will die Verkaufszahlen einmal je Monat für alle Filialen und einmal im Zeitverlauf für eine Filiale sehen. Dann will man Verkaufszahlen verschiedener Filialen oder verschiedener Monate miteinander vergleichen. Dazu braucht es aber nicht unterschiedliche Auswertungen, sondern es sind eigentlich nur verschiedene Sichten auf dieselben Daten respektive auf denselben Würfel oder Cube.

Wie aber erkläre ich dem Business einen Cube? Anhand eines Beispiels. Wir haben auf dem Internet ein einfaches Schulbeispiel eines Cubes - den Info.Cube - mit Verkaufszahlen erstellt. Dieser Info.Cube ist bewusst einfach gehalten und enthält nur die Grundfunktionalität.

Hier der Link: <http://democube.infokennel.ch/infocube.php?id=1>



	Beratung	Hardware	Software	Umsetzung	Total
2007	0	208	20	0	228
2008	3'423'057	3'525'854	1'485'422	888'991	9'323'323
2009	1'894'824	3'394'436	1'936'508	1'455'896	8'681'664
2010	2'163'637	2'727'514	1'714'951	2'593'262	9'199'364
2010Q1	682'037	827'500	469'782	388'604	2'367'923
2010Q2	496'170	660'252	409'437	473'727	2'039'587
2010Q3	499'160	656'419	425'597	799'328	2'380'504
2010Q4	486'269	583'344	410'134	931'604	2'411'351
2011	647'898	736'507	430'311	614'618	2'429'334
Total	8'129'415	10'384'519	5'567'212	5'552'767	29'633'913

Abbildung 1: Bildschirmkopie von Info.Cube

### Gewünschter Effekt des 1. Erziehungsschritts

Das Business sieht und lernt wie ein Cube funktioniert. Dabei lernen die Anwender gleichzeitig an einem konkreten Beispiel, was Dimensionen und Fakten sind und wie über Dimensionen die Fakten eingeschränkt werden können.

Ebenso lernen sie die Grundbegriffe Cube, Slice, Dice, Drill down und Drill in.

### Erziehungsschritt 2

Nachdem vom Cube her eine Vorstellung von Fakten und Dimensionen gegeben sind, sollen diese Begriffe nun vertieft werden. Als Einstieg zu diesem Thema frage ich normalerweise, wie die einzelnen Anwesenden den Begriff Umsatz definieren und wie dieser berechnet werden kann. Je nach Zusammensetzung der Gruppe ist die hiermit entfachte Diskussion kaum noch abzubrechen. Spätestens dann wird aber allen klar, dass die Definition Umsatz teilweise nicht einmal innerhalb einer Abteilung eindeutig ist, schon gar nicht über die Abteilungen hinweg. Hier wird auch sichtbar, warum es bei den verschiedenen Excel-Auswertungen teilweise Differenzen und Diskussionen gibt. Zeit also, auf die übereinstimmenden Dimensionen (conformed dimensions) und übereinstimmenden Fakten (conformed facts) einzugehen.

Dazu werden die Definitionen gemäss Kimball [Kim] folgendermassen erklärt:

#### Dimensionen

- Eine übereinstimmende Dimension (conformed dimension) ist eine Dimension, die mit verschiedenen Faktentabellen verknüpft werden kann.
- Das Definieren einer übereinstimmenden Dimension ist mehr eine politische oder fachliche Entscheidung und weniger eine technische. Diese Definition sollte daher von der höchsten Business-Stufe unterstützt sein.

#### Nutzen der übereinstimmenden Dimensionen

- Alle sprechen dieselbe Sprache und nutzen dieselben Dimensionen
- Übereinstimmende Dimensionen müssen zwingend zusammen mit allen betroffenen Bereichen des Business erarbeitet werden.
- Eine Konsenslösung muss gefunden werden, es sollten bewusst keine Spezialfälle abgebildet werden.

## Fakten

- Wenn es schwierig oder unmöglich ist, eine übereinstimmende Definition für ein Faktum zu finden, so müssen den unterschiedlichen Interpretationen auch unterschiedliche Namen gegeben werden

## Nutzen der übereinstimmenden Fakten

- Alle sprechen dieselbe Sprache und nutzen dieselbe Basis
- Daten unterschiedlicher Auswertungen sind vergleichbar
- Übereinstimmende Fakten müssen zwingend zusammen mit allen betroffenen Bereichen des Business erarbeitet werden.

## Fazit der übereinstimmenden Definitionen

- Modell ist einfach erweiterbar
- Alle arbeiten mit denselben Definitionen (Begriffe wie Netto oder Bruttoprämien werden zentral definiert und damit standardisiert).
- Alle arbeiten mit denselben Daten

Ein Data Warehouse, das auf übereinstimmenden Dimensionen und Fakten aufbaut, liefert konsistente Daten und kann als sogenannter „Single Point of Truth“ bezeichnet werden, also als die „eindeutige Wahrheit“.

## Gewünschter Effekt des 2. Erziehungsschritts

Das Business wird sich bewusst, dass eine abteilungsübergreifende Definition von Dimensionen und Fakten wichtig ist. Diese einheitliche Definition ist nicht für die Informatik, sondern vor allem für das Business und für ein einheitliches Reporting wichtig.

## **Erziehungsschritt 3**

Eine operative Datenbank wird oft mit Hilfe eines Entitäten-Relationen Modells entworfen. Viele Business-Anwender kennen diese Art der Modellierung. Versucht man aber, ein Data Warehouse mit dieser Methode zu modellieren, stösst man schnell an Grenzen. Wie soll die Dimension Zeit modelliert und verständlich dargestellt werden? Das Problem liegt bei den Hierarchien. Eine Dimension enthält meist eine oder gar mehrere Hierarchien. Damit dies verständlich dargestellt werden kann, verwenden wir die Notation ADAPT [ADAPT]. Diese Notation stellt die verschiedenen Hierarchiestufen mit ihren Abhängigkeiten dar. Versuchen wir gleich einmal die Dimension Zeit mit den allgemein bekannten Hierarchien darzustellen. Gehen wir davon aus, dass die kleinste Einheit (Granularität) ein Tag ist. Was ist dann die nächst höhere Stufe, auf die Fakten verdichtet werden können? Genau, entweder die Woche oder der Monat. Dabei wird nun klar, dass die Woche nicht in einen Monat verdichtet werden kann, denn leider gehört nicht jede Woche genau in einen Monat. Dafür kann ein Monat aber weiter zu einem Quartal, dieses zu einem Semester, dieses zu einem Jahr verdichtet werden.

So erhalten wir folgendes Modell für die Dimension Zeit.

Der Zusammenhang von Fakten und Dimensionen ist recht einfach. Hier auch ein Beispiel zum Datenmodell des Info.Cubes, der im Erziehungsschritt 1 vorgestellt wurde.

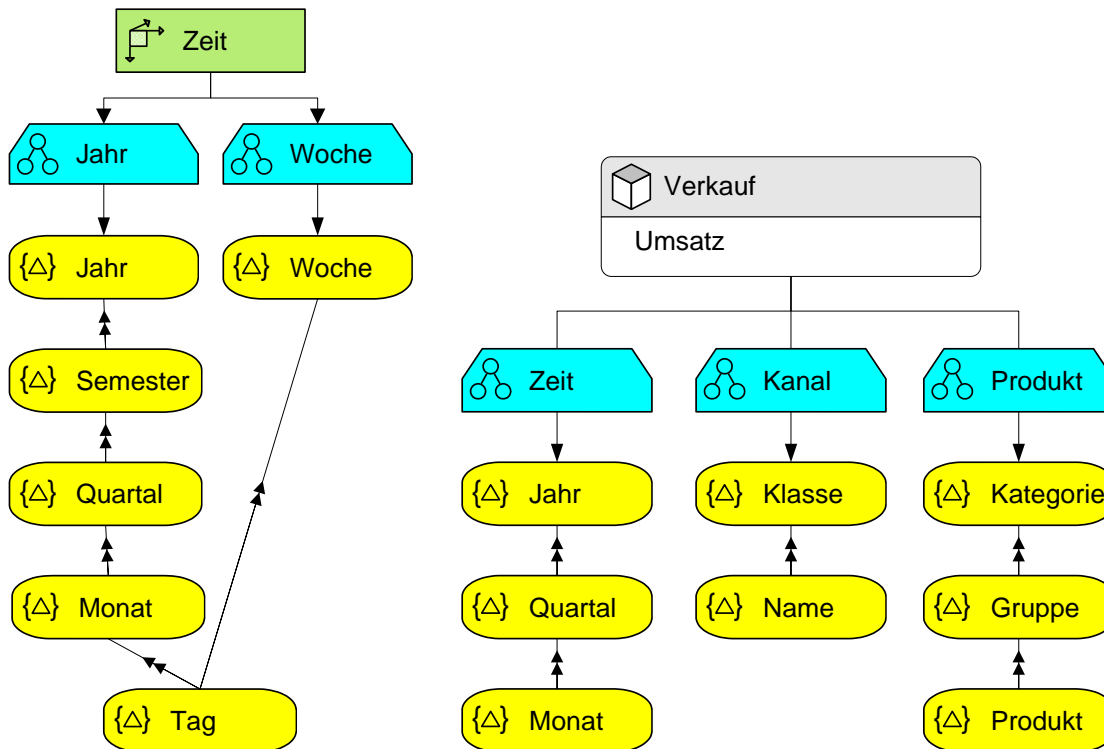


Abbildung 2: Die Dimension Zeit und Faktum Verkauf mit ADAPT dargestellt

Gewünschter Effekt des 3. Erziehungsschritts

Durch die Notation ADAPT wird dem Business gezeigt, dass Dimensionen verschiedene Hierarchiestufen enthalten können und wie diese einfach dargestellt werden können. Mit Hilfe der grafischen Notation können sich Informatik und Business finden und sprechen eine Art „gleiche Sprache“.

#### Erziehungsschritt 4

Bisher haben wir das Business, also unseren Kunden, soweit erzogen, dass wir uns zusammen über Fakten und Dimensionen mit Hierarchiestufen unterhalten können und hoffentlich auch dasselbe darunter verstehen.

Nun können wir die nächste Hürde angehen: die sogenannten „Slowly changing dimensions“, also Dimensionen, die sich über die Zeit ändern können. Auch hier ist es am Einfachsten, wenn wir mit einem Beispiel beginnen.

Im Info.Cube gibt es die beiden Produkte Gruppen Microsoft und Datenbanken. Gehen wir davon aus, dass wir bis Ende 2009 die Umsätze, die wir mit der Datenbank SQL-Server erzielt hatten, zu Microsoft gerechnet haben. Ab 2010 weisen wir diesen Umsatz aber unter Datenbanken aus.

Produktgruppe	Produkt	2009	2010
DB	Oracle	180	190
DB	Andere Datenbanken	50	50
MS(bis 2010)/DB(ab 2010)	SQL-Server	110	100
Datenbanken (Hierarchie 2009)		230	240
Datenbanken (Hierarchie 2010)		340	340

Abbildung 3: Auswirkung von Änderungen in einer Dimension

Nun stellt sich die Frage wie die Zahlen vor 2010 dargestellt werden sollen. Sollen diese so dargestellt werden, wie sie damals auch berechnet wurden, also in der Hierarchie von damals? Sollen die heutigen Zahlen auch in der Hierarchie von damals dargestellt werden? Sollen alle Zahlen in der Hierarchie dargestellt werden, die zum Zeitpunkt der Zahlen gültig waren? (das wären in der Grafik die blauen Felder). Oder sollen die Zahlen alle nach der heutige Hierarchie dargestellt werden? (so funktioniert der Info.Cube aktuell)

Anhand dieses einfachen Beispiels wird klar, dass diese Fragen für das Business wichtig sind, denn wenn die Daten alle in der aktuellen Hierarchie dargestellt werden, sind die alten Auswertungen nicht mehr gleich wie die heutigen. Werden aber alle Zahlen mit der Hierarchie ausgewiesen, die zum Zeitpunkt der Zahlen gültig war, dann gibt es Sprünge in der Zahlenreihe, die einen Vergleich erschweren oder verunmöglichen.

Falls die Frage nach der technischen Umsetzbarkeit nicht automatisch vom Business gestellt wird (meist kommt die Frage recht früh), so sollte diese doch noch kurz angesprochen werden. Konkret können die Dimensionen mit Versionen gespeichert werden.

PROD_ID	VALID_FROM	VALID_TO	PROD_NR	PROD_NAME	GROUP
111	01.01.1900	31.12.2009	2	SQL-Server	Microsoft
112	01.01.2010	31.12.9999	2	SQL-Server	DatenBanken
113	01.01.1900	31.12.9999	1	Oracle	DatenBanken

Abbildung 4: Speicherung einer Dimension mit Versionen

Wird eine Hierarchie so gespeichert, verweisen die Fakten immer auch auf die Hierarchie zum Zeitpunkt der Fakten. So ist eine Verdichtung nach der Hierarchie zu Zeitpunkt der Zahlen einfach. Diese versionierte Speicherung wird SCDII (Slowly Changing Dimension Typ II) genannt.

Braucht es nur die aktuelle Hierarchie, so müssen keine Versionen angelegt werden. Die Dimension kann einfach mit der aktuellen Hierarchie gespeichert werden. Diese Speicherung wird SCDI (Slowly Changing Dimension Typ I) genannt.

Gewünschter Effekt des 4. Erziehungsschritts

Das Business ist sich bewusst, was die Unterschiede zwischen SCDI und SCDII sind. Vor allem ist sich das Business bewusst, dass eine Lösung mit SCDII technisch etwas komplexer ist, diese Komplexität aber vor allem bei der Nachvollziehbarkeit der Business-Daten liegt.

Auswertungen über SCDII sind erklärungsbedürftig.. Wenn es möglich ist, bei Änderungen in den Dimensionen nur die heutige Hierarchie zu nutzen, so ist das klar einfacher.

### **Erziehungsschritt 5**

Der Erziehungsschritt 5 ist die Umsetzung des bisher Gelernten. Es geht darum, dass die Informatik gemeinsam mit dem Business die für Auswertungen benötigten Daten modelliert. Dabei ist es wichtig, dass nicht von der Quelle, also den im produktiven System vorhandenen Daten, sondern vom Auswertebedarfnis des Business ausgegangen wird. Ziel soll es sein ein BI-Lösung (Business Intelligence) zu bauen, die genutzt wird und nicht ein Data Warehouse, das verstaubt. Wie das Business intelligenter wird, weiss vor allem das Business; denn das Business wird intelligenter, wenn die Anwender die Daten und Informationen bekommen, die sie benötigen.

Gewünschter Effekt des 5. Erziehungsschritts

Unklarheiten zwischen Business und Informatik werden zu einem sehr frühen Zeitpunkt offensichtlich und können eliminiert werden. Das Business kennt das Bedürfnis nach Auswertungen, die Informatik kennt die Daten im produktiven System. So kann relativ einfach festgestellt werden, ob das produktive System die Basisdaten für die gewünschten Auswertungen überhaupt enthält, ob also alle „Messpunkte“ vorhanden sind. Wenn nicht, können die neuen Anforderungen an das produktive System aufgenommen oder weitere Datenquellen definiert werden.

## **Erziehungsschritt 6**

Nun sind Sie als Leser sicher verwundert, was denn das Business noch lernen könnte. Denn wenn das Datenmodell und die Quellen klar sind, folgt nur noch die Umsetzung. Das sollte nun doch wirklich eine Aufgabe der Informatik sein. Das ist auch so. Es stellt sich aber die Frage, wer in der Umsetzungsphase für die Projektleitung zuständig ist. Hier hat sich gezeigt, dass es hilfreich sein kann, wenn das Business für die Gesamtprojektleitung zuständig ist und die Informatik nur für die technische Teilprojektleitung.

Der Grund: Die Gesamtprojektleitung ist auch für die Projektfinanzen zuständig. So ist das Business an einem schnellen ROI (Return on Investment) interessiert. Daher wird die Projektleitung selten damit einverstanden sein, ein komplettes, grosses und komplexes Data Warehouse zu bauen, das dann in vielleicht 2 Jahren produktiv geht. Die Projektleitung wird eine Etappierung bevorzugen, die möglichst rasch erste Auswertungen mit dem Data Warehouse ermöglicht. Dies hat den grossen Vorteil, dass das Data Warehouse schrittweise entwickelt werden kann und Fehler, die immer vorkommen können, rasch erkannt und eliminiert werden können. Das Business, das sicher auch testen muss, ist mit den Tests der kleinen Teile auch weniger überfordert. Der erste Hunger nach Daten und Auswertungen mit dem Data Warehouse kann rasch gestillt werden und es kommt wieder Appetit nach neuen Auswertungen auf.

### **Gewünschter Effekt des 6. Erziehungsschritts**

Dank einer schrittweisen Entwicklung kann auch flexibel auf neue Bedürfnisse des Business eingegangen werden. Denn oft ändern sich die Bedürfnisse und auch die Prioritäten. So können neue Fakten oder Dimensionen, die plötzlich dringend werden, relativ rasch in das bestehende Data Warehouse eingebaut werden.

Bei diesem Erziehungsschritt kann man sich fragen, ob hier noch die Informatik das Business erzieht oder ob das Business eher die Informatik zu agilen Lösungen erzieht.

## **Fazit**

Durch die aufgeführten „Erziehungsschritte“ lernen die Anwender folgende Punkte:

- Wie kann mit einem Datenwürfel gearbeitet werden.
- Was sind übereinstimmende Dimensionen und Fakten, wieso ist eine übereinstimmende Definition von Fakten und Dimensionen wichtig.
- Wie kann ein multidimensionales Modell und vor allem Hierarchien mit ADAPT dargestellt werden.
- Was sind Slowly Changing Dimensions.

Es ist sinnvoll, wenn beim Vorstellen dieser Grundlagen nicht nur die Anwender, sondern auch die Informatik dabei ist. Denn nicht immer sind alle Beteiligten der Informatik auf dem gleichen Wissensstand oder verwenden alle die gleichen Fachbegriffe.

Dadurch, dass die Anwender und die Informatik nun die gleichen Begriffe verwenden und wissen, was diese bedeuten, wird die Zusammenarbeit einiges einfacher. Die Anwender können abschätzen, was einfach und was komplex ist. Daher kann es vorkommen, dass die Anwender den Informatikern nicht mehr so einfach glauben, wenn diese behaupten, gewisse Anforderungen seien nicht umsetzbar. Der Vorteil, dass die Anwender auch die Komplexität sehen, überwiegt aber klar. So kommt es häufig vor, dass die Anwender zwar gerne eine Dimension als SCDII hätten, wegen des Aufwandes aber darauf verzichten. Diese Einbindung der Anwender in die Definition der Anforderungen in Form von Fakten und Dimensionen ist zentral für das Gelingen eines Data Warehouse Projektes.

Ein weiterer Erfolgsfaktor ist die bewusste schrittweise Definition der Anforderungen und deren schrittweise Implementation. Dies ist einfacher möglich, wenn die Gesamtprojektleitung dem Business zugeteilt ist und die Informatik nur für die technische Umsetzung zuständig ist.

Es ist klar, dass zum Grundwissen im Bereich Data Warehousing noch viel mehr dazu gehört, als in diesem Artikel beschrieben. So ist eine Architektur, die mit dem Projekt mitwächst, klare Aufteilungen der Aufgaben und Kompetenzen, die Wahl der richtigen Hilfsmittel bis hin zur Wahl der Hard- und Software wichtig für das Gelingen eines Projektes. Doch nützt ein technisch optimal und perfekt aufgebautes Data Warehouse nichts, wenn es nicht vom Business akzeptiert und genutzt wird. Die Mitarbeit und ein Grundverständnis beim Business ist zwar keine Garantie, dass das Projekt gelingt, aber eine zentrale Voraussetzung.

Kontaktadresse:

Dr. Andrea Kennel  
InfoPunkt Kennel GmbH  
Bahnhofstr. 48  
CH-8600 Dübendorf

Telefon: +41 (0) 44 820 71 40  
E-Mail [andrea@infokennel.ch](mailto:andrea@infokennel.ch)  
Internet: [www.infokennel.ch](http://www.infokennel.ch)

### **Literatur**

[Kim] The Data Warehouse Toolkit, R. Kimball. Wiley Computer Publishing, ISBN 0-471-15337-0

[ADAPT] Bulos/Forsman: Getting Started with ADAPT™ - OLAP Database Design, Whitepaper, 2006 ([http://www.symcorp.com/downloads/ADAPT\\_white\\_paper.pdf](http://www.symcorp.com/downloads/ADAPT_white_paper.pdf))