

Einführung in Datenbanken

Übung 14: ER-Modell, Logischer Entwurf

Prof. Dr. Stefan Brass

PD Dr. Alexander Hinneburg

Martin-Luther-Universität Halle-Wittenberg

Wintersemester 2023/24

<http://www.informatik.uni-halle.de/~brass/db23/>

Präsenzaufgabe: Kardinalitäten (2)

- Richtig war Lösung C:



- Ein Kunde kann mehrere Bestellungen aufgeben.
- Ein Kunde muss keine Bestellungen aufgeben.
Die DB enthält auch „Kunden“, die nur einen Katalog bestellt haben.
- Es gibt keine Sammelbestellungen, d.h. jede Bestellung ist von höchstens einem Kunden.
- Es gibt keine anonymen Bestellungen, d.h. jede Bestellung ist von mindestens einem Kunden.

Zusammen also ist jede Bestellung von genau einem Kunden.

Logischer Entwurf (3)

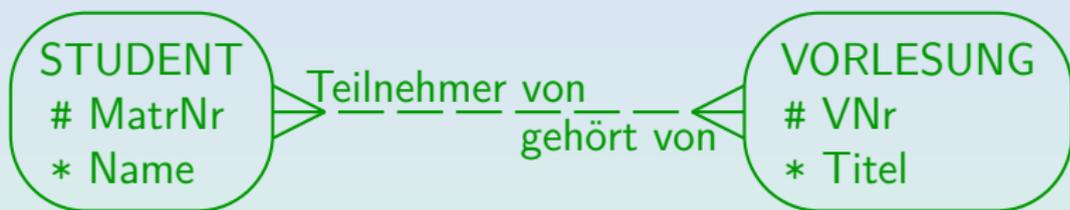
- Beispiel: Eins-zu-viele Beziehung wird in Fremdschlüssel auf der „viele“ Seite übersetzt:



- $DOZENTEN(\underline{Name}, Tel)$
 $VORLESUNGEN(\underline{Nr}, Titel, Name \rightarrow DOZENTEN)$
- Falls das Relationship auf der Vorlesungs-Seite gestrichelt ist, erlaubt der Fremdschlüssel Nullwerte.
- Nicht alle Kardinalitäten lassen sich übersetzen.
- Manchmal sind zusätzliche Integritätsbedingungen nötig.

Logischer Entwurf (6)

- Viele-zu-viele Beziehungen werden in Tabellen mit aus zwei Fremdschlüsseln zusammengesetzten Schlüssel übersetzt:



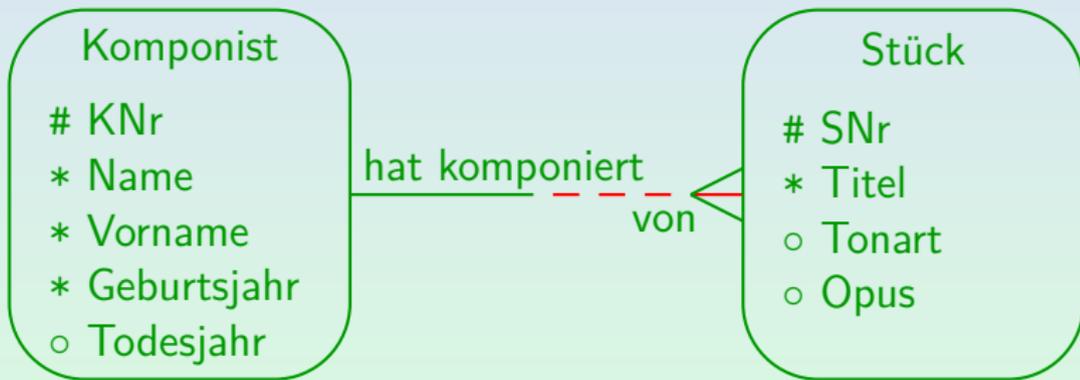
- STUDENTEN(MatrNr, Name)
VORLESUNGEN(VNr, Titel)
TEILNAHME(MatrNr→STUDENTEN, VNr→VORLESUNGEN)

- Studierende müssen so keine Vorlesungen hören, und Vorlesungen müssen keine Teilnehmer haben.

Man kann in die Tabellen STUDENTEN und VORLESUNGEN Zeilen einfügen, und die Tabelle TEILNAHME leer lassen.

Hausaufgabe 13.1 (6)

- Es soll aber auch möglich sein, Stücke mit unbekanntem Komponisten (d.h. **ohne Komponist**) in die Datenbank zu speichern.



- Gestrichelte Linie auf der „Stück“-Seite:
Stücke müssen nicht unbedingt an der Beziehung teilnehmen.

Hausaufgabe 13.1 (8)

- Alte Notation (Chen-Notation mit (min,max)-Kardinalitäten):



- Die Attribute sind hier weggelassen (wären in Ovalen außerhalb der Rechtecke für die Entity-Typen).
- Die Intervalle spezifizieren, wieviel ausgehende Kanten ein Entity des jeweiligen Typs haben kann. Aber Vorsicht:
 - In UML-Klassendiagrammen genau umgekehrt.
 - In Barker-Notation ist der Krähfuß für die maximale Kardinalität * auf der anderen Seite.

Die minimalen Kardinalitäten (erste Komponente) entsprechen dagegen gestrichelter (0) bzw. durchgezogener Linie (1) auf der gleichen Seite.

Hausaufgabe 13.1 (10)

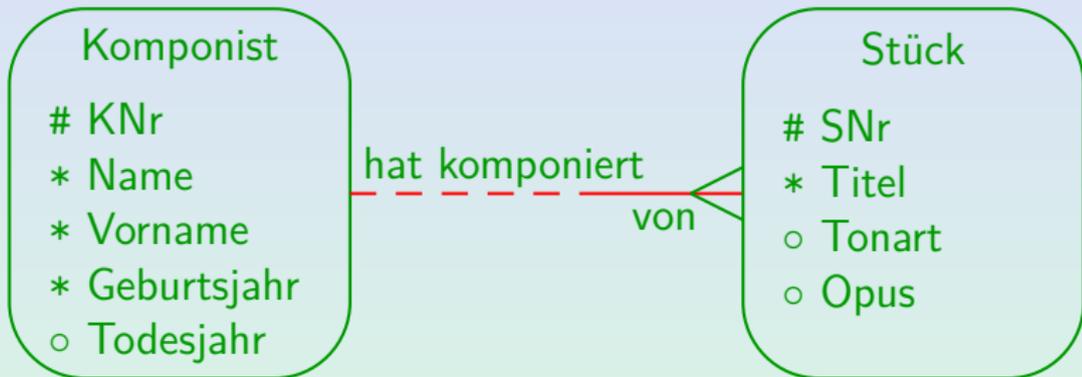
- Womit haben Sie das Diagramm gezeichnet?
 - A. Handschriftlich
 - B. \LaTeX (pur oder mit oder mit TikZ)
 - C. Graphviz
 - D. Oracle SQL Developer Data Modeler
 - E. Anderes Zeichenprogramm
- Welches andere Zeichenprogramm?
 - A. [draw.io] → [https://app.diagrams.net/]
 - B. yEd: [https://www.yworks.com/products/yed]
 - C. Dia: [http://dia-installer.de/]
 - D. Powerpoint, Libre Office, Word, ...
 - E. Sonstiges (welches?)

Hausaufgabe 13.1 (11)

- Wieviel Zeit hat das Zeichnen des ER-Diagramms gebraucht?
 - A. 0–15 min.
 - B. 15–30 min
 - C. 30–60 min.
 - D. Mehr als 60 min.
- Können Sie ein Zeichenprogramm besonders empfehlen?
- Möchten Sie von einem Zeichenprogramm abraten?

Hausaufgabe 13.1: Beispiele für Fehler (1)

- Gestrichelte Linie und durchgezogene Linie vertauscht:



- Es kann so auch Komponisten in der Datenbank geben, die noch kein einziges Stück komponiert haben.

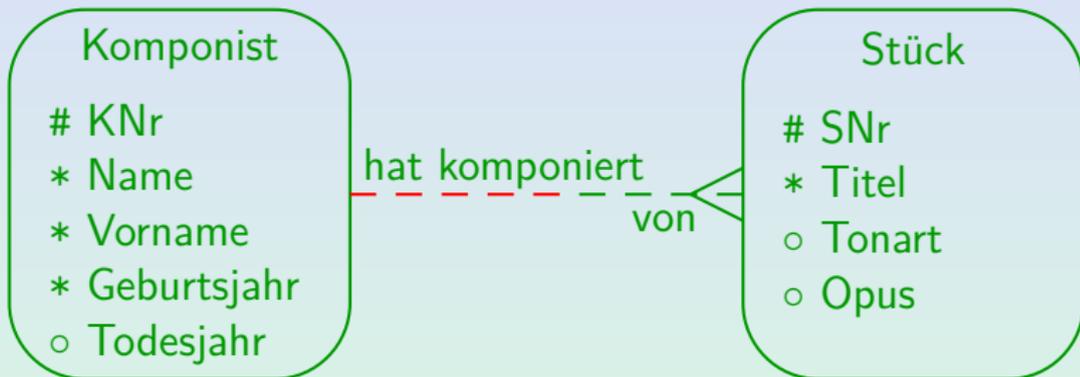
Das könnte übergangsweise sinnvoll sein, widerspricht aber der Aufgabenstellung.

- Stücke müssen so immer einen Komponisten haben.

In den Beispiel-Daten gibt es schon 10 Stücke mit einem Nullwert in `knr`.

Hausaufgabe 13.1: Beispiele für Fehler (2)

- Line ganz gestrichelt:

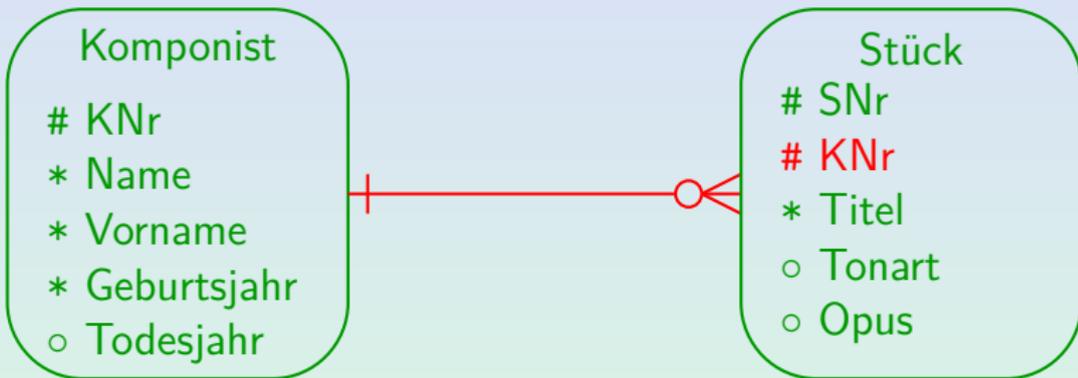


- Schema ist zu allgemein:
 - Alle verlangten Zustände können dargestellt werden.
 - Aber auch einige Zustände, die nach der Aufgabenstellung verboten sind (Komponisten ohne Stücke).

Sie können fragen, ob es nicht vielleicht doch möglich sein sollte, das Komponisten zunächst ohne Stücke in die DB eingefügt werden.

Hausaufgabe 13.1: Beispiele für Fehler (4)

- Falsche Notation bzw. falsches Zeichenprogramm:



- Dies ist nicht die Barker-Notation für das Relationship.
Wenn Sie ein Vorlesungsskript etc. haben, das diese Notation erklärt, würde mich das interessieren.
- Der Fremdschlüssel in Stück ist für ein ER-Diagramm falsch.
Er ist auch redundant zum dargestellten Relationship.

Hausaufgabe 13.1: Beispiele für Fehler (5)

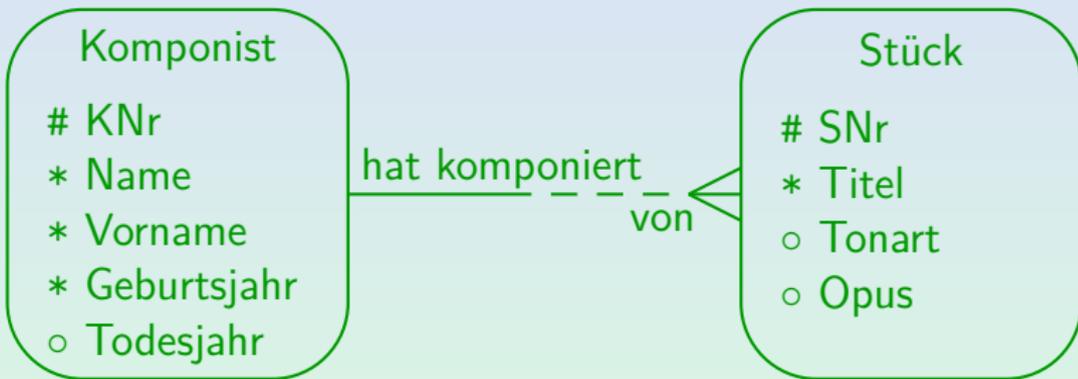
Verschiedenes:

- Ellipsen statt „Softboxes“: Rechtecke mit abgerundeten Ecken.
- Markierung „○“ für optionale Attribute fehlt.
- Spalten für Schlüssel und Namen nicht ausgerichtet.
 - Die Symbole #, * und ○ müssen übereinander stehen.
- Die zwei Entity-Typen sollten durch eine horizontale Linie verbunden sein (ggf. auch vertikal). Nicht schräg oder abgeknickt.

Bei komplexeren Diagrammen sind Relationship-Kanten mit einem Knick möglich (Wechsel zwischen horizontal und vertikal). Schräge Kanten sind in seltenen Fällen auch möglich, wenn sie für das Verständnis der Struktur von Vorteil sind. Normalerweise sollte man sich aber auf horizontale und vertikale Linien beschränken. Und Überkreuzungen minimieren.

Exkurs: Logischer Entwurf (1)

- Als Motivation war gesagt, dass das ER-Schema den Tabellen **Komponist** und **Stueck** entsprechen sollte.

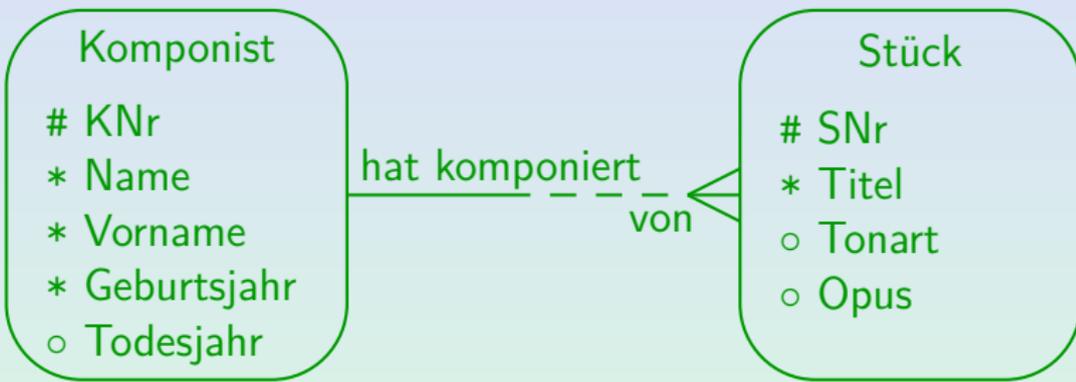


- Im „Logischen Entwurf“ (zweite Phase des DB-Entwurfs) wird das ER-Schema ins relationale Modell übersetzt.

Ziel ist, dass die relationalen Zustände genau den Zuständen zum ER-Schema entsprechen. Wenn der konzeptionelle Entwurf richtig gemacht wurde, entsprechen diese wiederum den Situation des relevanten Welt-Ausschnitts.

Exkurs: Logischer Entwurf (2)

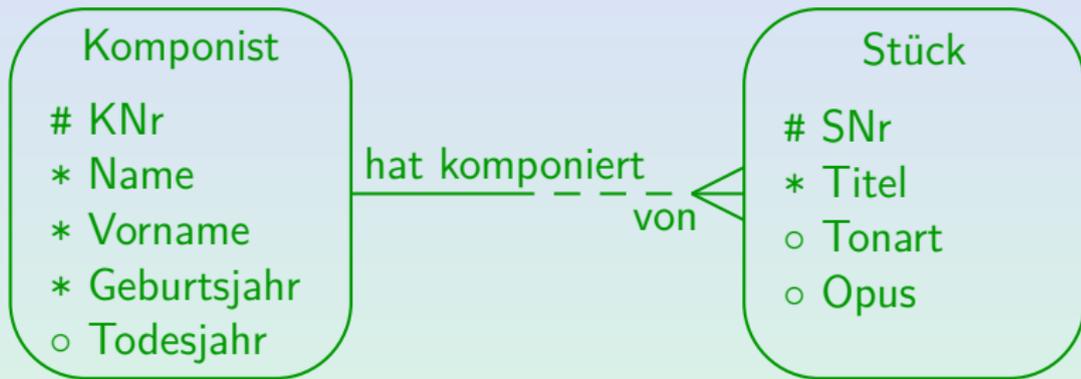
- ER-Diagramm (Eingabe des logischen Entwurfs):



- Man legt zunächst für jeden Entity-Typ eine entsprechende Tabelle an und übernimmt Attribute, Optionalität, Schlüssel:
 - `Komponist(KNr, Name, Vorname, Geburtsjahr, Todesjahro)`
 - `Stueck(SNr, Titel, Tonarto, Opuso)`

Exkurs: Logischer Entwurf (3)

- ER-Diagramm:

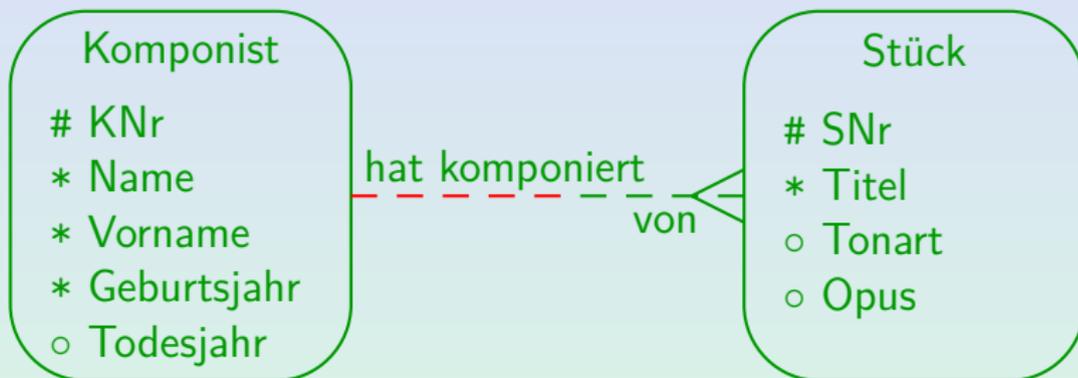


- Eins-zu-viele (1:n) Beziehungen (Kähenfuß nur auf einer Seite) werden übersetzt, indem man den Schlüssel der „Eins“-Seite als Fremdschlüssel der „Viele“-Seite hinzufügt:

- `Stueck(SNr, Titel, Tonart○, Opus○)`
`KNr○ → Komponist)`

Exkurs: Logischer Entwurf (5)

- Tatsächlich hat man nur dieses ER-Diagramm übersetzt:



- Beim erzeugten relationalen Schema ist es möglich, Komponisten einzutragen, die von keinem Stück referenziert werden.

Immerhin können alle vom Original-Diagramm erlaubten Zustände auch als Zustände zum relationalen Schema dargestellt werden. Aber das relationale Schema erlaubt Zustände, die keine Entsprechung im ER-Diagramm haben.

Exkurs: Logischer Entwurf (6)

- Lösung: Zusätzliche allgemeine Integritätsbedingung.
 - In natürlicher Sprache: „Zu jedem Komponisten muss es mindestens ein Stück mit seiner KNr geben.“

Oder: Es darf keine Komponisten ohne Stücke geben.

- Als logische Formel:

$$\forall \text{Komponist } k: \exists \text{Stueck } s: s.KNr = k.KNr$$

- Als SQL-Anfrage, die Fehlermeldungen berechnet:

```
SELECT 'Komponist ohne Stücke: ' || Name ||  
      ', ' || Vorname AS errmsg  
FROM   Komponist k  
WHERE  NOT EXISTS (SELECT *  
                  FROM   Stueck s  
                  WHERE  s.KNr = k.KNr)
```

Hausaufgabe 13.2 (1)

- Zeichnen Sie ein ER-Diagramm in der Notation der Vorlesung (Barker-Notation),
- Sie müssen das ER-Diagramm im PDF-Format abgeben.
- Mit handgezeichneten (und eingescannten) Diagrammen können Sie maximal 10 Punkte erreichen (von 11).

Es können Punkte auch für schlechtes Aussehen abgezogen werden.

- Das Relationship muss wie in der Vorlesung gezeigt beschriftet werden, die Position der beiden Namen kann aber anders sein.

Wenn es gar nicht anders geht, reicht auch ein Name in der Mitte.

Hausaufgabe 13.2 (2)

- Über den **Stoff** soll der **Bezeichner** und falls vorhanden die **Tagesdosis** und deren **Einheit** gespeichert werden. Beides kann fehlen. Ein Stoff wird durch den Bezeichner identifiziert.

Stoff

- # Bezeichner
- Tagesdosis
- Einheit

Man sollte wohl eine Integritätsbedingung hinzufügen, dass Tagesdosis und Einheit nur zusammen Null oder nicht Null sein können. Das wurde aber hier nicht erwartet. Wenn man es wirklich als Teil des Schemas der Vitamin-DB sieht (hier ist das nur Motivation), müsste die Einheit nicht Null sein, weil sie immer für die Menge gebraucht wird.

Hausaufgabe 13.2 (3)

- Eine randomisiert kontrollierte, klinische Studie hat einen Digital Object Identifier (DOI), der eine Veröffentlichung identifiziert. Weiterhin werden noch das Jahr der Veröffentlichung und der Titel gespeichert, die immer für eine gespeicherte Studie bekannt sein müssen.



Ich vermute, dass Studie und Veröffentlichung mehr oder weniger Synonyme sind (in einer Veröffentlichung wird über eine Studie berichtet). Sie hätten fragen können.

Hausaufgabe 13.2 (4)

- Ein(e) Autor:in einer Studie wird durch die ORCID-ID identifiziert.

Die Non-Profit-Organisation ORCID (Open Researcher Contributor Identification Initiative) wurde 2010 gegründet und vergibt seit 2012 ORCID-IDs.

- Zu einer Autorin oder einem Autor wird in jedem Fall der Name gespeichert. Falls die Autorin oder der Autor einen Titel und eine Organisation angegeben hat, für die sie oder er arbeitet, werden diese ebenfalls gespeichert.

Autor:in

ORCID-ID

* Name

○ Titel

○ Organisation

Hausaufgabe 13.2 (5)

Hinweis/Korrektur:

- Die Anmerkung zur ORCID-Organisation sollte eigentlich keinen Einfluss auf die Lösung haben.

Falls Sie später an wissenschaftlichen Arbeiten beteiligt sein sollten, oder einen Doktor machen, wäre es wahrscheinlich gut, sich eine ORCID-ID geben zu lassen, und die bei Ihren Arbeiten mit abzuspeichern, wo das möglich ist. Man kann dann später in Literatur-Datenbanken leichter „Ihre“ Arbeiten finden, selbst wenn es noch andere Autoren mit dem gleichen Namen gibt, oder Ihre Vornamen unterschiedlich abgekürzt werden.

- Tatsächlich muss man sich aber Fragen, was ist mit Autoren, die keine ORCID-ID haben?

Auch vor 2012 sind schon Studien erschienen, und nicht jeder Autor registriert sich bei dem ORCID-Verein.

Hausaufgabe 13.2 (6)

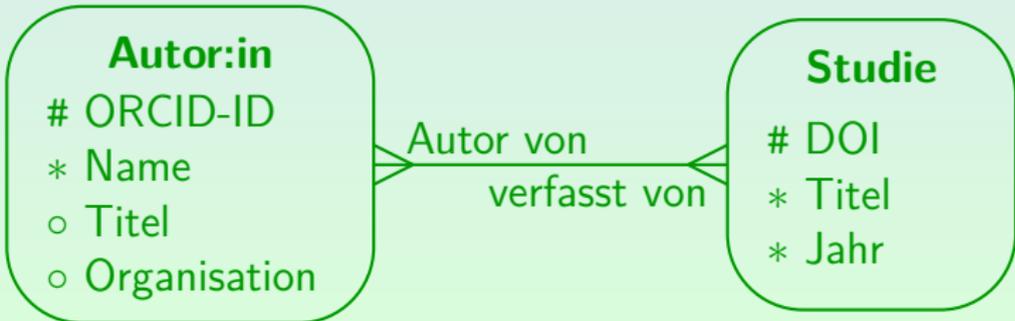
Hinweis/Korrektur, Forts.:

- Als Datenbank-Entwerfer sollten Sie nachfragen, wenn Sie erkennen, dass die Anforderungen problematisch/unklar sind.
- Es war nicht so geplant, aber es wäre wahrscheinlich richtig, Autoren eine datenbank-interne ID als Schlüssel zu geben, und die ORCID-ID zusätzlich zu speichern, soweit bekannt oder vorhanden.
- Man hätte dann die Integritätsbedingung, dass es keine zwei Autoren/Autorinnen geben kann, bei denen die ORCID-ID definiert und gleich ist.

Es kann dagegen beliebig viele Autoren/Autorinnen geben, bei denen die ORCID-ID einen Nullwert enthält. Bei Oracle und PostgreSQL sind UNIQUE-Schlüssel so definiert. Andere Systeme (MS SQL Server) würden nur eine Zeile mit einem Nullwert erlauben.

Hausaufgabe 13.2 (7)

- Ein(e) Autor:in ist mindestens an einer Studie in der Datenbank beteiligt, sonst würde sie oder er nicht gespeichert werden. Ein(e) Autor:in kann jedoch auch an mehreren Studien beteiligt sein.
- Eine Studie hat mindestens eine(n) Autor:in. Oft wurde eine Studie jedoch von mehreren Autorinnen und Autoren geschrieben.



Hausaufgabe 13.2 (8)

Exkurs zum DB-Entwurf:

- Eine viele-zu-viele-Beziehung mit verpflichtender Teilnahme auf beiden Seiten ist praktisch etwas schwierig.

Wenn die Spezifikation aber so ist, müssen Sie es so modellieren.

Sie könnten den Kunden vielleicht nochmal fragen, ob er es wirklich so will.

- Man muss neue Autoren zusammen mit ihrer ersten Studie erfassen.

Bei existierenden Autoren reicht es dagegen, die ORCID-ID einzugeben.

Vielleicht gibt man immer alle Daten ein, und das System meldet sich, wenn es einen Konflikt gibt (gleiche ORCID-ID mit unterschiedlichem Namen).

- Im Oracle Designer Entwurfswerkzeug war dieser Typ von Relationships nicht unterstützt, weil das in der Praxis nicht vorkommt ...

Hausaufgabe 13.2 (9)

- Zu einem Stoff kann es noch keine Studie geben.
- Zu einem Stoff können auch mehrere Studien veröffentlicht worden sein.
- Wir nehmen hier an, dass eine Studie genau einen Stoff untersucht.

