

1986

# SERIE RESEARCH MEMORANDA

REVALUATIE VAN HET E-R MODEL

E.R.K Spoor & S.J.L. Kramer

Researchmemorandum 1986-25



VRIJE UNIVERSITEIT  
FACULTEIT DER ECONOMISCHE WETENSCHAPPEN  
A M S T E R D A M



**REVALUATIE VAN HET E-R MODEL**  
**Opbouwende kritiek op tien jaar Entity-Relationship Model**

*E. R. K. Spoor & S. J. L. Kramer*

De verscheidenheid aan definities en (grafische) representatievormen die, een decennium na de geboorte van het Entity-Relationship (E-R) Model, het oeuvre van de E-R cultuur kenmerken, rechtvaardigen een herbezinning op de fundamenten en de opbouw van het model. In dit artikel worden de belangrijkste ontwikkelingen sinds 1976 kritisch en mathematisch beschouwd teneinde een aanzet te geven tot een consistent en vooral bruikbaar model.



## REVALUATIE VAN HET E-R MODEL

### Opbouwende kritiek op tien jaar Entity-Relationship Model

*E. R. K. Spoor & S. J. L. Kramer*

#### 1. Inleiding

Het vervaardigen van een dataschema als grondslag voor een database georiënteerd informatiesysteem kan, afgezien van voorafgaande fasen (onder andere acquisitie), ruwweg worden verdeeld in twee stappen:

1. ontwerp van een conceptueel dataschema;
2. ontwerp van een fysiek dataschema.

De eerste stap is een abstractieproces waarin gegevens worden onderscheiden en geordend die relevant zijn voor de omgeving waarin de database dient te functioneren. De tweede stap kan worden omschreven als een transformatieproces waarin het conceptuele schema dient te worden omgezet in een schema dat wordt beperkt door de keuze van een (commerciële beschikbaar) databank management systeem. Deze systemen ondersteunen geheel of gedeeltelijk de inmiddels als klassiek bestempelde hiërarchische, netwerk en relationele modellen.<sup>1</sup>

Er zijn tenminste twee redenen te bedenken om een conceptueel dataschema te ontwerpen alvorens de beperkingen van een der fysieke modellen in rekening te brengen.

1. De onafhankelijkheid van het conceptuele model ten opzichte van de fysieke modellen staat een wijziging van modelkeuze op fysiek niveau toe.
2. Het conceptuele model weerspiegelt een zuiverder beeld van het doelgebied dan de aan technische beperkingen onderhevige fysieke modellen.

Tien jaar geleden schreef Chen een artikel over entiteiten en relaties,<sup>2</sup> basisconcepten voor het ontwerpen van een conceptueel dataschema. Dit fundament fungeert inmiddels als verdichtingspunt van een indrukwekkende verzameling artikelen waarin uitbreidingen op het oorspronkelijke idee worden geïntroduceerd, kritieken worden geleverd, mathematische onderbouwingen en praktische toepassingen worden weergegeven.

Door de verscheidenheid aan uitbreidingen en de verschillen in definities is er van een consistent geheel nauwelijks sprake.<sup>3</sup> Zonder de pretentie te hebben volledig te zijn, belichten we in dit artikel de belangrijkste ontwikkelingen in de Entity-Relationship-cultuur en beschouwen uitbreidingen van het oorspronkelijke model kritisch, teneinde door keuze en consistente benadering te komen tot een bruikbaar, goed gedefinieerd model, het Uitgebreide-Entiteit-Relatie (U-E-R) model. Het axiomatische fundament van dit model wordt gevormd door het onderscheid tussen een individuele entiteit en het type van deze entiteit, waarmee eigenschappen, verzamelingen en klassen van entiteiten een nieuw aanzien krijgen.

In hoofdstuk 2 wordt de basis gelegd voor de begrippen entiteit, relatie en attribuut. De uitbreidingen op deze basis die vervolgens kritisch worden beschouwd zijn bestaansafhankelijkheid (hoofdstuk 3), klassificatie van relatietypen (hoofdstuk 4) en abstracties zoals aggregatie en generalisatie (hoofdstuk 5). De theoretische beschouwingen worden praktisch toegelicht met voorbeelden ontleend aan het bedrijfsspel VUMAS.<sup>4</sup>

## 2. Entiteiten, relaties, attributen

Een *entiteit* is een fenomeen dat uniek kan worden geïdentificeerd en bestaansrecht heeft in een abstractieproces.<sup>5</sup> Het subjectieve karakter van een abstractieproces verhindert een nauwkeuriger definitie, zodat met voorbeelden als een specifiek persoon, voorwerp, concept of gebeurtenis het begrip entiteit moet worden verduidelijkt.

De klassifikatie van entiteiten binnen een organisatorische context krijgt een formele basis door de functie

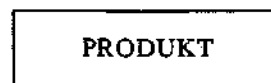
$$\text{TYPE: } \{e \mid e \text{ is een entiteit}\} \rightarrow N,$$

waarbij  $N$  een verzameling namen aangeeft. Twee willekeurige entiteiten  $e_1$  en  $e_2$  behoren tot hetzelfde type als  $\text{TYPE}(e_1) = \text{TYPE}(e_2)$ . De functie  $\text{TYPE}$  voegt, anders gezegd, aan iedere entiteit een naam toe die aangeeft tot welke klasse de entiteit behoort.<sup>6</sup> Als  $E$  een type-naam is, dan geven we met  $\delta E$  de klasse van entiteiten van het type  $E$  aan. Een *entiteitverzameling* is een groep entiteiten van hetzelfde type.

Ten behoeve van de eenvoud en duidelijkheid zal in het vervolg van dit artikel geen onderscheid worden gemaakt tussen  $\delta E$  als klasse van entiteiten en  $\delta E$  als verzameling, tenzij anders aangegeven.<sup>7</sup>

Voorbeelden van entiteitstypen zijn **PERSOON**, **WERKNEMER** en **PRODUKT**. Passend bij het laatstgenoemde type, bestaat er een entiteitverzameling  $\delta \text{PRODUKT}$ . Deze bevat entiteiten zoals "Abyco" en "Etheliet" waarvoor geldt dat  $\text{TYPE}(\text{"Abyco"}) = \text{TYPE}(\text{"Etheliet"}) = \text{PRODUKT}$ . Zij zijn derhalve van hetzelfde type.

In het E-R diagram heeft een entiteitverzameling de vorm van een rechthoek met daarin de naam van de verzameling.



Bij een entiteit horen eigenschappen die kenmerkend zijn voor het type waartoe de entiteit behoort. Een eigenschap (of *attribuut*<sup>8</sup>) voegt aan een entiteit een waarde toe uit een verzameling waarden van eenzelfde type, aangeduid door een predikaat. Voorbeelden van waardenverzamelingen zijn **PLAATSEN**, **ACHTERNAMEN**, **VOORNAMEN** en **JAREN** met respectievelijk "Columbus OHIO", "Nicklaus", "Jack" en "1940" als een overeenkomstige waarde uit elke verzameling.

Zij *attribuut*  $A_j$  ( $1 \leq j \leq n$ ) zo'n beschrijvende eigenschap, waarvan elke waarde wordt geassocieerd met een individuele entiteit.  $A_j$  is een functie die een bepaalde entiteitverzameling  $\delta E$  afbeeldt op een waardenverzameling  $V_j$  of cartesisch produkt van waardenverzamelingen als volgt:

$$A_j: \delta E \rightarrow V_j \quad \text{of} \quad A_j: \delta E \rightarrow V_{j_1} \times V_{j_2} \times \dots \times V_{j_n}.$$

Deze definitie volgend, beeldt een attribuut **WOONPLAATS** een **PERSOON**-entiteit af op de waardenverzameling **PLAATSEN**, zoals bijvoorbeeld **WOONPLAATS** ("Nicklaus") = "Columbus/OHIO". Een voorbeeld van een cartesisch produkt van verzamelingen is het attribuut **NAAM**, welke een afbeelding geeft naar het produkt van de verzamelingen **VOORNAMEN** en **ACHTERNAMEN**. De namen van de waardenverzamelingen kunnen overeenkomen met die van attributen hoewel het twee verschillende concepten betreft.

Een entiteitverzameling van het type  $E$  kan nu formeel worden gekarakteriseerd in de vorm

$$E(A_1/V_1, A_2/V_2, \dots, A_n/V_n)$$

waarbij elk stel  $A_j/V_j$  een *attribuut-waardenverzameling-paar* voorstelt.<sup>9</sup> Een entiteitverzameling wordt dus aangeduid door zijn type en zijn attribuut-waardenverzameling-paren. Een *entiteit*  $e$  uit  $\delta E$  krijgt betekenis door het tupel  $(v_1, v_2, \dots, v_n)$  met  $v_j = A_j(e)$ .<sup>10</sup> Zo wordt met ("Nicklaus", "Columbus OHIO") een entiteit uit de entiteitverzameling **WERKNEMER** (**NAAM/ACHTERNAMEN**, **PLAATS/PLAATSEN**) aangegeven. Hoewel hier twee attribuut-waardenverzameling-paren zijn aangegeven, worden individuele entiteiten veelal slechts genoemd met hun bepalende eigenschap(pen) — in dit geval een **NAAM**-waarden. Met de naam "Nicklaus" refereren we derhalve aan een unieke entiteit met al zijn attribuut-waarden. Het attribuut **NAAM** heeft zo een identifikatie-functie, doch bij de gratie van de exclusiviteit van de **NAAM**-waarde. Indien er meerdere werknemers met dezelfde achternaam bestaan kan deze niet meer worden gebruikt ter identifikatie. Er dient dan een combinatie van attributen te worden gevonden. Een alternatieve keuze is het introduceren van een kunstmatig attribuut, zoals een **PERSOONSNUMMER**, met een bijbehorende waardenverzameling om de entiteiten uniek te kunnen onderscheiden.

Het attribuut dat, of de kollektie van attributen die de entiteitverzameling één-éénduidig afbeeldt op de korresponderende kollektie waardenverzamelingen heet de *entiteit-sleutel*.<sup>11</sup> Als  $\{A_1/V_1, \dots, A_s/V_s\}$  een

deelverzameling is van  $\{A_1/V_1, \dots, A_n/V_n\}$  dan heet zo'n kollektie identificerende attributen,  $I = \{A_1, \dots, A_s\}$  ( $1 \leq s \leq n$ ), een sleutel van de entiteitverzameling  $\delta E$  dan en slechts dan als  $I$  een minimale verzameling van attributen is, die een één-éénduidige afbeelding geeft van  $E$  naar het cartesisch produkt  $V_1 \times \dots \times V_s$ , zodanig dat geen enkele echte deelverzameling van  $I$  ook een sleutel van  $E$  is. In het geval dat er meerdere kandidaat-sleutels voorkomen, moet er een afweging worden gemaakt tussen efficiëntie en semantische betekenis, teneinde er één te kiezen tot *primaire sleutel* (PS). We kunnen bijvoorbeeld schrijven dat PS (WERKNEMER) = {PERSOONSNUMMER}. In de karakterisering van een entiteitverzameling laten we, bij gelijke benamingen, de waardenverzamelingen weg en onderstrepen de primaire sleutel: WERKNEMER (PERSOONSNUMMER, NAAM, PLAATS). Deze notatie zal in het vervolg van dit artikel domineren.

Tussen entiteiten bestaan associaties. Bijvoorbeeld de mengverhouding tussen twee produkten. Een ander voorbeeld is de associatie, betiteld als WERKT\_IN, waarmee een WERKNEMER-entiteit wordt verbonden met de PLOEG-entiteit waarin hij werkt. Een associatie tussen entiteiten noemen we voortaan een *relatie*.<sup>12</sup> Er zijn verschillende typen relaties tussen verschillende typen entiteiten mogelijk waarbij we de relaties van hetzelfde type groeperen in een *relatieverzameling*.<sup>13</sup> Formeel betekent dit een uitbreiding van de eerder genoemde TYPE-functie en wel als volgt:

$$\text{TYPE} : \{e \mid e \text{ is een entiteit}\} \cup \{r \mid r \text{ is een relatie}\} \rightarrow N$$

met de additionele voorwaarden

- 1) Als  $\text{TYPE}(r_1) = \text{TYPE}(r_2)$  dan zijn  $r_1$  en  $r_2$  van hetzelfde relatietype en zijn  $r_1$  en  $r_2$  gedefinieerd op dezelfde verzameling entiteiten.
- 2)  $\forall e, r : \text{TYPE}(e) \neq \text{TYPE}(r)$ .

De verzameling namen  $N$  bevat nu zowel namen van entiteitverzamelingen als van relatieverzamelingen. Zij  $R$  de naam van een relatieverzameling, dan wordt met  $\delta R$  de klasse van alle relaties van het type  $R$  aangegeven. Een relatie  $r$  uit  $\delta R$  krijgt pas betekenis door z'n eigenschapswaarden in combinatie met de gerelateerde entiteiten. Deze entiteiten fungeren als identifikatie van de voornoemde eigenschapswaarden. Elke relatieverzameling wordt aldus gekarakteriseerd door z'n naam, z'n attributen en de betrokken entiteitverzamelingen en komt overeen met een wiskundige relatie tussen  $m$  entiteiten<sup>14</sup> waaraan bovendien bepaalde eigenschappen zijn verbonden. We noteren dit als:

$$R(P_1/E_1, \dots, P_m/E_m; A_1/V_1, \dots, A_n/V_n).^{15}$$

Hierbij representeert  $P_i$  de rol die  $\delta E_i$  heeft in  $\delta R$ , zoals GRONDSTOF en HALFPRODUKT in een MENGVERHOUDING de rollen vertolken van de entiteitverzameling  $\delta \text{PRODUKT}$ . Dit voorbeeld toont al aan dat de  $\delta E_i$  niet noodzakelijk onderling disjunkt behoeven te zijn. Evenmin is er een beperking van het aantal entiteitverzamelingen waarop een relatieverzameling is gedefinieerd.

Evenals entiteiten kunnen relaties eigenschappen bezitten door middel van attribuut-waardenverzamelingparen  $A_j/V_j$  met  $j = 1, 2, \dots, n$ . Een attribuut  $A_j$  beeldt in dit geval een relatieverzameling  $\delta R$  af op een waardenverzameling (of het cartesisch produkt van waardenverzamelingen)  $V_j$ :

$$A_j : \delta R \rightarrow V_j \quad \text{of} \quad A_j : \delta R \rightarrow V_{j_1} \times V_{j_2} \times \dots \times V_{j_n}$$

Een attribuut wordt slechts dan aan een relatieverzameling toegerekend als het niet past bij één van de gerelateerde entiteitverzamelingen, doch van alle betrokkenen afhangt. BENODIGDE\_HOEVEELHEID is een attribuut van de relatieverzameling  $\delta \text{MENGVERHOUDING}$  dat geen inhoudelijke betekenis heeft bij alleen een PRODUKT-entiteit maar dat wordt bepaald door de relatie tussen beide.  $\delta \text{MENGVERHOUDING}$  wordt gekarakteriseerd door MENGVERHOUDING (GRONDSTOF/PRODUKT, HALFPRODUKT/PRODUKT; BENODIGDE\_HOEVEELHEID/GEWICHTEN). Een ander voorbeeld van een relatieverzameling maar dan tussen verschillende entiteitverzamelingen is WERKT\_IN (WERKNEMER, PLOEG:). Omdat rolnaam en entiteittype gelijk zijn, is in deze karakterisering een redundante notatie vermeden. Beide relatietypen worden hieronder afgebeeld met behulp van een ruit welke de entiteitverzamelingen met elkaar verbindt.



Figuur 1.

In het E-R diagram is tevens een nadere karakterisering van de relatietypen aangegeven.<sup>16</sup> De aanduidingen 1 en M in figuur 1 bepalen het aantal entiteiten dat is toegestaan in een relatieverzameling. Voor een binaire relatieverzameling tussen entiteitverzamelingen  $\delta E_1$  en  $\delta E_2$  kennen we in het E-R diagram drie soorten aanduidingen:

- 1 : 1      Een entiteit van type  $E_1$  is gerelateerd aan ten hoogste één entiteit van type  $E_2$ .
- 1 : M      Een entiteit van type  $E_1$  is gerelateerd aan een willekeurig aantal entiteiten van type  $E_2$ , maar iedere entiteit van type  $E_2$  is gerelateerd aan ten hoogste één entiteit van type  $E_1$ .
- N : M      Een entiteit van type  $E_1$  is gerelateerd aan een willekeurig aantal entiteiten van type  $E_2$  en omgekeerd.

Deze aanduidingen worden genoteerd op de verbindinglijnen aan de zijden van de betreffende entiteitverzamelingen.

De tot hertoe gedefinieerde begrippen vormen de basis van het originele E-R model. Het is mogelijk om daarmee in een E-R diagram een (deel)organisatie vrij overzichtelijk te representeren. Echter, de kracht van een model schuilt in de semantische uitdrukingskracht en op dat aspect valt er nog wel wat te verbeteren.<sup>17</sup> De aanduiding van (meervoudige) afhankelijkheid van entiteiten is zo'n punt, evenals de klassifikatie van bepaalde eigenschappen van relaties. Het toevoegen of juist onderdrukken van details en het modelleren van afwijkende visies zijn eveneens enige zaken die nog niet aan de orde zijn gekomen, maar die wel tot het U-E-R model dienen te behoren. De diagrammatische representatie wordt vanzelfsprekend aangepast overeenkomstig deze nieuwe begrippen. Wij zullen hieronder allereerst afhankelijkheid van entiteiten bespreken, gevolgd door klassifikatie van relaties. Als laatste komen abstracties aan bod waaronder we meerdere begrippen verstaan zoals aggregatie, generalisatie en deelverzameling hiërarchieën en de notatie van attributen.



### 3. Bestaansafhankelijkheid

Het bestaan van een entiteit kan soms afhankelijk zijn van de aanwezigheid van andere entiteiten.<sup>18</sup> Ter illustratie beschouwen we de entiteitstypen EINDPRODUKT en PRIJSINFO en het relatietype AANBOD (figuur 2).



Figuur 2. PRIJSINFO is bestaansafhankelijk van EINDPRODUKT.

De informatie over aanbiedingsprijzen is niet altijd relevant als we over eigenschappen van een eindprodukt praten en wordt daarom ondergebracht in een apart entiteitstype PRIJSINFO. Entiteiten hieruit hebben echter alleen bestaansrecht als de bijbehorende EINDPRODUKT-entiteit bestaat. Anders geformuleerd is het bestaan van prijsinformatie afhankelijk van het bestaan van het eindprodukt: PRIJSINFO is bestaansafhankelijk van EINDPRODUKT via het relatietype AANBOD. Het diagram geeft dit aan door een pijl die gericht is naar en zich bevindt aan de zijde van de afhankelijke entiteitverzameling.<sup>19</sup>

Een entiteitverzameling  $\delta E$  heet *bestaansafhankelijk* van de entiteitverzamelingen  $\delta E_1, \dots, \delta E_k$  in een relatieverzameling gekarakteriseerd door  $R(E_1, \dots, E_k, E; A_1, \dots, A_n)$  als het bestaan van een entiteit  $e$  uit  $\delta E$  afhankelijk is van het bestaan van een relatie  $(e_1, \dots, e_k, e)$ . Het  $k$ -tupel entiteiten  $(e_1, \dots, e_k)$  (met  $e_i \neq e$  en  $e_i \in \delta E_i$  voor  $i = 1, \dots, k$ ) is element van de verzameling  $S^e$  met  $S^e$  de *existentie-deelverzameling* van  $e$  met betrekking tot  $\delta E_1, \dots, \delta E_k$ .  $S^e \subseteq \delta E_1 \times \dots \times \delta E_k$  en bevat die tupels van entiteiten die bepalend zijn voor het bestaan van  $e$ . Een relatieverzameling welke geen onderwerp is van bestaansafhankelijkheid heet *regulier*.

De verzameling  $S^{Abyco}$  waarbij "Abyco" element is van de entiteitverzameling  $\delta$ PRIJSINFO, bestaat uit de entiteit "Abyco" uit de entiteitverzameling  $\delta$ EINDPRODUKT. Dit voorbeeld toont aan dat er nog een reden kan zijn waarom een entiteitverzameling bestaansafhankelijk kan zijn. De entiteiten van het type PRIJSINFO worden namelijk geïdentificeerd door de sleutel van de EINDPRODUKT-entiteiten (NAAM) en kunnen alleen zinvol bestaan zolang de EINDPRODUKT-entiteit aanwezig is. De beschrijving van een afhankelijke entiteitverzameling bevat derhalve ook de sleutelattributen van de bepalende entiteitverzameling. PRIJSINFO kan er dan bijvoorbeeld uitzien als PRIJSINFO (NAAM, MAX\_AANB\_PRIJS, MIN\_AANB\_PRIJS). Indien entiteiten niet kunnen worden geïdentificeerd door hun eigen attributen maar slechts via een relatieverzameling dan wordt dit een *zwakke relatieverzameling* genoemd en heet de desbetreffende entiteitverzameling een *zwakke entiteitverzameling*. De entiteiten hieruit worden (mede) geïdentificeerd door een *externe sleutel* in tegenstelling tot de gebruikelijke *interne sleutel*.<sup>20</sup>

Een ander voorbeeld wordt geïllustreerd door figuur 3. We kunnen alleen een ploeg formeren als er (geschikte) werknemers voor te vinden zijn. In U-E-R terminologie zeggen we dat een PLOEG-entiteit alleen in de databank kan worden opgenomen als er WERKNEMER-entiteiten zijn te vinden die met deze PLOEG-entiteit kunnen worden geassocieerd; PLOEG is dus bestaansafhankelijk van WERKNEMER.

In dit tweede voorbeeld is, evenals in het eerste, een bestaansafhankelijke entiteitverzameling aanwezig maar in dit geval is deze slechts bestaansafhankelijk uit semantische overwegingen. Zowel een WERKNEMER- als een PLOEG-entiteit wordt geïdentificeerd door eigen sleutelattribuut-waarden. De notatie is voor beide vormen gelijk hoewel een zwakke relatieverzameling op een lager conceptueel niveau nog enige extra aanduidingen krijgt. In paragraaf 5.1 komen we hierop terug. Met bovenstaande definitie en notatie kunnen binnen een relatieverzameling ook meerdere bestaansafhankelijkheden voorkomen en kan een entiteitverzameling eveneens meer dan één keer bestaansafhankelijk zijn van verschillende andere entiteitverzamelingen. Het is zelfs mogelijk dat twee entiteitverzamelingen wederzijds bestaansafhankelijk zijn.<sup>21</sup>



Figuur 3. PLOEG bestaansafhankelijk van WERKNEMER.

#### 4. Klassifikatie van relatietypen

In het oorspronkelijke E-R model wordt een relatietype slechts nader aangeduid door kardinaliteitskenmerken.<sup>16</sup> In die zin heeft het E-R model beperkte uitdrukkingmogelijkheden. Latere uitbreidingen onderkennen een nadere karakterisering van relatietypen terzake de limiet, de beperking en de volledigheid van deelname van entiteiten in een relatie. Het U-E-R model onderscheidt drie klassen van eigenschappen:<sup>22</sup>

- 1 Kardinaliteits-klasse
- 2 Restriktie-klasse
- 3 Participatie-klasse

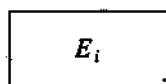
##### 4.1. Kardinaliteits-klasse

Zoals reeds eerder vermeld, kent het E-R model drie aanduidingen voor de *kardinaliteits-klasse*<sup>16</sup>, te weten 1:1, 1:M (of M:1) en N:M - welke aangeven wat het *maximum* aantal malen is, dat een entiteit uit een entiteitverzameling wordt *toegestaan* om in een relatieverzameling voor te komen. Over het verplichte minimum aantal wordt hier geen uitspraak gedaan. Hoewel deze notatievorm goed funktioneert, is niettemin een overstap naar een eenvoudiger en modernere manier van noteren te prefereren.<sup>23</sup> Een voorbeeld hiervan is te zien in figuur 4. Elke werknemer mag slechts in één ploeg werken maar in een ploeg mogen wel meerdere verschillende werknemers werken (de situatie M:1).

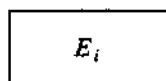


Figuur 4. Voorbeeld van een M : 1 relatieverzameling met een nieuwe notatie.

De kardinaliteits-klasse van een relatieverzameling gekarakteriseerd door  $R(E_1, \dots, E_m; A_1, \dots, A_n)$  tussen  $m$  entiteitverzamelingen  $\delta E_i$  ( $1 \leq i \leq m$ ) kan als volgt worden vastgelegd:



Elke entiteit  $e_i$  uit  $\delta E_i$  mag in niet meer dan één relatie uit  $\delta R$  voorkomen.



Elke entiteit  $e_i$  uit  $\delta E_i$  mag onbeperkt in meerdere relaties uit  $\delta R$  voorkomen.

##### 4.2. Restriktie-klasse

Naast een limiet op het aantal relaties waarin een entiteit betrokken mag zijn (de kardinaliteits-klasse), kan er ook een restrictie bestaan aangaande *welke* entiteiten met elkaar een relatie mogen vormen.<sup>24</sup>

Een relatieverzameling gekarakteriseerd door  $R(E_1, \dots, E_m, E; A_1, \dots, A_n)$  is *beperkt* indien elke entiteit  $e$  uit  $\delta E$  slechts mag zijn gerelateerd aan die entiteiten  $(e_1, \dots, e_m)$  waarvoor geldt dat  $(e_1, \dots, e_m)$  behoort tot  $C^e$ , de *restriktie-deelverzameling* van  $e$ , waarbij  $C^e \subset \delta E_1 \times \dots \times \delta E_m$ . De elementen uit  $C^e$  zijn de enige waarmee  $e$  een relatie mag vormen.

Als een relatieverzameling beperkt is geven we dit aan in het U-E-R diagram met een keper onderaan de ruit zoals in figuur 5.a. Een ketel verwerkt geen grondstoffen maar is beperkt tot halfprodukten (een deelverzameling van produkten). In het geval dat E bestaansafhankelijk is kan het voorkomen dat de entiteiten uit  $\delta E$  slechts zijn gerelateerd aan die entiteiten waarvan hun bestaan afhangt.<sup>25</sup> Dit is een situatie waarin  $C^e = S^e$ , de restriktie-deelverzameling van  $e$  is gelijk aan de existentie-deelverzameling van  $e$  voor  $\forall e \in \delta E$ . De relatieverzameling die beperkt is en een bestaansafhankelijkheid aangeeft ten opzichte van dezelfde entiteitverzameling heet *bestaansbeperkt* en ziet eruit als in figuur 5.b. Een keper en een pijl geeft aan dat elke ploeg alleen bestaat zolang erin één of meer werknemers werken. Omdat een ploeg slechts in één productiefase kan worden ingeschakeld, wordt de samenstelling ervan beperkt tot die werknemers welke geschikt zijn om in die betreffende fase te werken.

Een relatieverzameling zonder restrictie heet *onbeperkt*.



Figuur 5. a) Een beperkte relatieverzameling.

b) Een bestaansbeperkte relatieverzameling.

### 4.3. Participatie-klasse

Zoals het model tot hiertoe is gedefinieerd, kan een U-E-R diagram weergeven hoe vaak een entiteit mag voorkomen in relaties en met welke entiteiten dit mag gebeuren. Hiermee kan nog niet worden aangegeven in welke mate een entiteitverzameling verplicht participeert in een relatieverzameling, of wel wat het *noodzakelijke* minimum aantal verschijningen van een entiteit in een relatieverzameling is. Deze participatie van entiteiten in een relatieverzameling openbaart zich in drie vormen: partieel, totaal en compleet.<sup>26</sup>

Zij  $R$  de naam van een relatieverzameling gekarakteriseerd door  $R(E_1, \dots, E_m; A_1, \dots, A_n)$  en zij  $\delta E_k$  ( $1 \leq k \leq m$ ) een willekeurige entiteitverzameling. Een relatieverzameling wordt gedefinieerd als

totaal ten opzichte van  $E_k$  als  $\forall e_k \in \delta E_k: \exists (e_1, \dots, e_m) \in C^k: r = (e_1, \dots, e_k, \dots, e_m) \in \delta R$ , ofwel elke entiteit van  $\delta E_k$  moet minimaal één keer voorkomen in de (beperkte)  $\delta R$ .

compleet ten opzichte van  $E_k$  als  $\forall e_k \in \delta E_k: \forall (e_1, \dots, e_m) \in C^k: r = (e_1, \dots, e_k, \dots, e_m) \in \delta R$ , ofwel elke entiteit van  $\delta E_k$  komt verplicht voor in alle relaties van de (beperkte)  $\delta R$ .

partieel ten opzichte van  $\delta E_k$  in alle andere gevallen.

Uit praktische overwegingen beperken we de participatie-klasse tot de vormen *totaal* en *partieel*.<sup>27</sup> In het algemeen zal een relatieverzameling partieel zijn, wat wordt aangegeven met een dwarsliggend streepje op de verbindingslijn met de relatie-ruit naast de entiteitverzameling ten opzichte waarvan de relatieverzameling partieel is. Een relatieverzameling die totaal is krijgt geen extra grafisch kenmerk. Voorbeelden hiervan vinden we in figuur 6 waar KETEL partieel participeert in WERKT\_MET en PLOEG totaal participeert omdat met elke PLOEG-entiteit minimaal één KETEL-entiteit moet zijn verbonden.



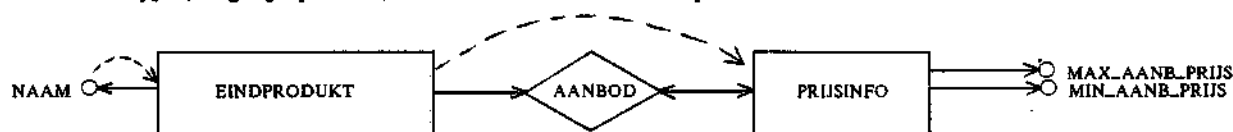
Figuur 6. Participatie-klasse: WERKT\_MET

## 5. Abstracties

Naarmate een deelorganisatie groter of een toepassing complexer is, zal het lastiger worden om in het conceptuele schema wijzigingen aan te brengen en zal de behoefte groeien om bepaalde aspecten nader te kunnen specificeren of generaliseren. Daarnaast moet het mogelijk zijn om voor verschillende doelgebieden afzonderlijke submodellen te definiëren en deze in een andere meer of minder gedetailleerde vorm te kunnen presenteren. Bovendien is het belangrijk om te allen tijde de overzichtelijkheid van het U-E-R diagram te handhaven. Om dit te realiseren definiëren we hieronder recente uitbreidingen van het E-R model onder de noemer *abstracties*. Onder abstractie verstaan we het onderdrukken van bepaalde details die voor de betreffende toepassing of gebruikers niet relevant zijn.<sup>29</sup> Bij het ontwerpen kunnen verschillende conceptuele niveaus van detaillering ontstaan die afhankelijk zijn van de omvang en complexiteit van de schema's en de mate van behoefte aan details. Structuurprimitieven ten behoeve van deze ontwerpactiviteiten zijn *aggregatie*,<sup>30</sup> waarbij een relatie wordt getransformeerd tot een hoger niveau object, en *hiërarchieën*.<sup>31</sup> Deze laatste vorm van abstractie voorziet in een gecontroleerde manier om relevante details pas naderhand te introduceren met behulp van *deelverzameling of generalisatie hiërarchieën*.

### 5.1. Attributen

Het U-E-R model zoals tot hertoe gedefinieerd, staat op een "hoger" conceptueel abstractie-niveau, gevormd uit entiteit- en relatietypen. Het predikaat "hoger" refereert hier aan de mate van details in het U-E-R diagram. Uitbreiding van de weergave met bijvoorbeeld attribuut-aanduidingen "verlaagt" het abstractie-niveau.<sup>32</sup> Een attribuut wordt aangegeven door een gerichte lijn vanuit de entiteit- of relatieverzameling naar een cirkeltje welke een attribuut-waardenverzameling-paar representeert.<sup>33</sup> Het kan ook voorkomen dat een attribuut binnen één entiteit meerdere waarden heeft.<sup>34</sup> In dat geval noteren we twee pijltjes op de lijn naar het attribuut. Geheel anders is een attribuut waarvan de waarde als *optioneel* wordt gespecificeerd.<sup>35</sup> Analoog aan de participatie-klasse van relatietypen, wordt dit aangegeven met een dwars-streepje (vergelijk partieel) als een attribuut-waarde optioneel is.



Figuur 7. Een relatieverzameling op een lager conceptueel niveau (met attributen).

Indien een attribuut ook de rol van identificerende sleutel heeft, dan beelden we dit af met een onderbroken lijn die weer terug loopt naar de korresponderende entiteitverzameling. We zien dit in figuur 7, waar elke EINDPRODUKT-entiteit wordt geïdentificeerd door de eigen produktnaam. Een PRIJSINFO-entiteit wordt echter bepaald door de naam van het overeenkomstige EINDPRODUKT-entiteit. Dit is het eerder genoemde voorbeeld van een bestaansafhankelijke entiteitverzameling met externe sleutel.<sup>20</sup> Aangezien de sleutelattributen van het type EINDPRODUKT de entiteiten van PRIJSINFO bepalen loopt de identificerende stippellijn vanuit EINDPRODUKT naar het afhankelijke type PRIJSINFO. Figuur 8 is een voorbeeld van een attribuut (BENODIGDE\_HOEVEELHEID) dat aan een relatieverzameling toebehoort.



Figuur 8. Attribuut van een relatieverzameling.

### 5.2. Aggregatie

Bij de konstruktie van een relatieverzameling, is het niet uitgesloten dat wat in de ene situatie als een relatieverzameling wordt gezien, in een andere situatie in dezelfde deelorganisatie als een zelfstandig begrip wordt gehanteerd. Er is bijvoorbeeld een heel verschil tussen de combinatie van een ketel, een bepaalde hoeveelheid eenheden produktietijd, een eindprodukt, etcetera enerzijds en een dit alles omvattende naam

als PRODUKTIE-II anderzijds. Op een lager conceptueel niveau kan een dergelijk relatietype met bijbehorende entiteitverzamelingen en attributen uitgebreid worden gerepresenteerd. In vele gevallen echter, vooral wanneer men een groter geheel beschouwt, zijn al deze details irrelevant en is het beter om over het concept PRODUKTIE-II te spreken. Het is bijgevolg wenselijk een relatieverzameling met de betrokken entiteitverzamelingen te abstraheren tot een hoger niveau object, een *aggregaat*,<sup>30</sup> dat in het U-E-R model funktioneert als een nieuwe entiteitverzameling en in die vorm weer met andere entiteitverzamelingen in relatie kan staan. We zullen hieronder formeel aantonen dat een relatie kan worden beschouwd als een entiteit en zullen hiertoe een nieuwe entiteitverzameling konstrueren.

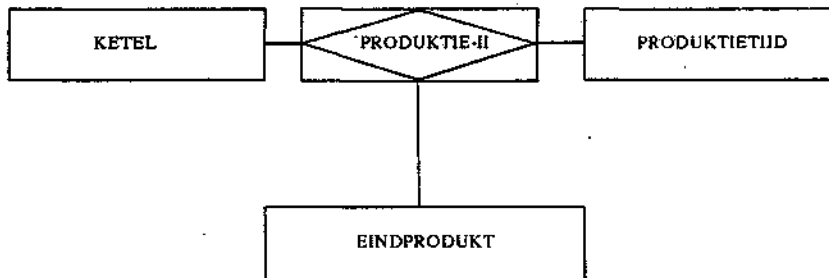
Zij  $R(E_1, \dots, E_m; A_1, \dots, A_n)$  de karakteristiek van een relatieverzameling die is gedefinieerd op de entiteitverzamelingen  $\delta E_1, \dots, \delta E_m$ . We willen elk element  $r$  uit  $\delta R$  beschouwen als een entiteit in de entiteitverzameling  $\delta R^*$  waarbij  $\text{TYPE}(r) = R^*$ , een naam die element is van de verzameling type namen  $N$ . Voor deze overgang van relatie naar entiteit bestaat een bijektieve funktie  $\phi: \delta R \rightarrow \delta R^*$ , met  $\phi(r) = r^*$ . Met behulp hiervan kan in twee stappen uit een relatie  $r \in \delta R$  een entiteit  $r^* \in \delta R^*$  worden gekonstrueerd. Ten eerste formeren we nieuwe sleutelattributen  $A_k^i$  gevolgd door een transformatie van de overige (niet-sleutel)attributen  $A_j$ .

Zij  $\{A_1^i, \dots, A_{s_i}^i\}$  de sleutel van  $\delta E_i$  ( $1 \leq i \leq m$ ).

- (1)  $\forall A_k^i \in \delta R$  ( $1 \leq i \leq m, 1 \leq k \leq s_i$ ) definieer  $A_k^{*i}$  zodanig dat  
 $\forall r \in \delta R, \exists r^* \in \delta R^* : \phi(r) = r^* \wedge A_k^i(r) = A_k^{*i}(r^*)$ .
- (2)  $\forall A_j \in \delta R$  ( $1 \leq j \leq n$ ) definieer  $A_j^*$  zodanig dat  
 $\forall r \in \delta R, \exists r^* \in \delta R^* : \phi(r) = r^* \wedge A_j(r) = A_j^*(r^*)$ .

Daar de nieuwe attributen  $A_k^{*i}$  en  $A_j^*$  uitbreidingen zijn van respectievelijk  $A_k^i$  en  $A_j$ , kunnen zonder beperking der algemeenheid de oorspronkelijke attribuutnamen zonder ster worden gebruikt.

We noteren de nieuwe entiteitklasse  $\delta R^*$  dan als  $R^*(A^1, \dots, A^m, A_1, \dots, A_n)$  waarbij  $R^*$  de naam van het aggregaat is,  $\{A^1, \dots, A^m\}$  de sleutel vormt (elke  $A^i = \{A_1^i, \dots, A_{s_i}^i\}$  is de sleutel van  $\delta E_i$  ( $1 \leq i \leq m$ )) en  $A_1, \dots, A_n$  de niet-sleutelattributen van het aggregaat zijn.



Figuur 9. Aggregatie van de relatieverzameling  $\delta$ PRODUKTIE-II.

Het is nu mogelijk een relatieverzameling te aggregeren en te behandelen als een entiteitverzameling. Grafisch geven we dit weer door een rechthoek om de ruit van de geaggregeerde relatieverzameling (figuur 9).<sup>36</sup> De relatieverzameling met karakteristiek PRODUKTIE-II (KETEL, PRODUKTIETIJD, EINDPRODUKT; HOEVEELHEID) is hier geaggregeerd tot entiteitverzameling PRODUKTIE-II\* (KETELNUMMER, PRODUKTIETIJD\_EENHEID, PRODUKTNAAM, HOEVEELHEID).

### 5.3. Deelverzameling hiërarchie

Er bestaan situaties waarin binnen een entiteitverzameling bepaalde deelverzamelingen kunnen worden onderscheiden waarvoor afwijkende eigenschappen gelden. Een voorbeeld: Sommige eindprodukten worden niet alleen op de thuishmarkt verhandeld, maar kunnen ook op de exportmarkt worden verkocht. Ze ondergaan op beide markten een andere behandeling zodat ook enkele relevante eigenschappen voor beide verschillend zijn. De entiteitverzameling EINDPRODUKT bestaat in dit geval uit twee, onderling niet-

disjunkte deelverzamelingen, gekarakteriseerd door:

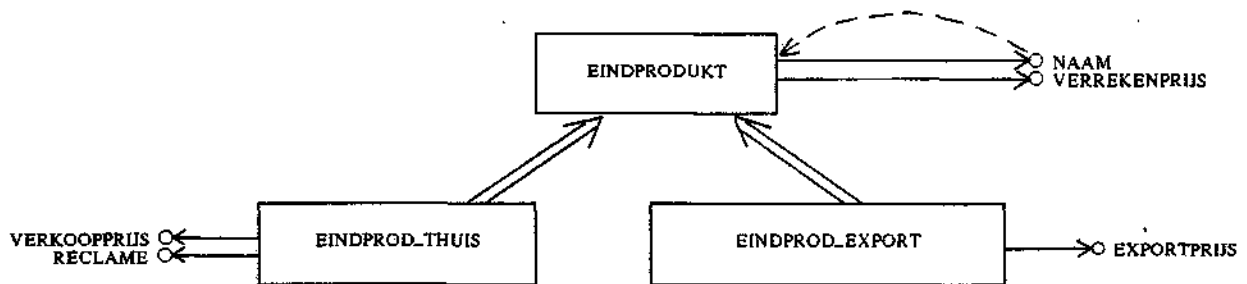
EINDPROD\_THUIS (NAAM, VERREKENPRIJS, VERKOOPPRIJS, RECLAME) en  
EINDPROD\_EXPORT (NAAM, VERREKENPRIJS, EXPORTPRIJS).

De attributen NAAM en VERREKENPRIJS zijn hierin herkenbaar als gemeenschappelijke kenmerken van een eindprodukt: EINDPRODUKT (NAAM, VERREKENPRIJS). In figuur 10 zijn de deelverzamelingen  $\delta E_{\text{EINDPROD\_THUIS}}$  en  $\delta E_{\text{EINDPROD\_EXPORT}}$  met hun eigen kenmerken ondergebracht op een gedetailleerd niveau in de hiërarchie.<sup>37</sup> Dezelfde deelverzameling hiërarchie maar dan in een ruimere context is terug te vinden in figuur 11.

Formeel is entiteitverzameling  $\delta F$ , gekarakteriseerd door  $F(A_1, \dots, A_s, A_{s+1}, \dots, A_n, A_{n+1}, \dots, A_p)$ , *deelverzameling* van de *superverzameling*  $\delta E$ , gekarakteriseerd door  $E(A_1, \dots, A_s, A_{s+1}, \dots, A_n)$ , als elke entiteit  $f$  uit  $\delta F$  kan worden afgeleid uit  $\delta E$ , ofwel

$$\forall f \in \delta F: \exists e \in \delta E: f = (v_1, \dots, v_n, w_{n+1}, \dots, w_p) = e = (v_1, \dots, v_n).^{38}$$

Alle attributen die op  $\delta E$  zijn gedefinieerd moeten ook op  $\delta F$  zijn gedefinieerd. Bovendien mag de deelverzameling  $\delta F$  additionele attributen bezitten die niet bij  $\delta E$  horen.



Figuur 10. Deelverzameling hiërarchie.

#### 5.4. Generalisatie hiërarchie

Het tweede soort hiërarchie staat bekend onder de naam generalisatie (de tegenhanger van specialisatie).<sup>39</sup> Het verschil tussen deelverzameling en generalisatie hiërarchie zit in het feit dat elke entiteit  $e \in \delta E$  in het eerste geval in meerdere deelverzamelingen  $\delta E_i$  ( $1 \leq i \leq m$ ) mag voorkomen maar in een generalisatie hiërarchie slechts in ten hoogste één van deze deelverzamelingen. De generalisatie hiërarchie kan daardoor beter een *exklusieve* generalisatie worden genoemd. Zowel bij deelverzameling als bij generalisatie hiërarchie erven de entiteitverzamelingen op een lager abstractie-niveau alle attributen van het hogere niveau. Bovendien kunnen op het lagere niveau nog additionele attributen aanwezig zijn.<sup>40</sup>

Zij  $E_i(A_1, \dots, A_s, A_{s+1}, \dots, A_n, A_{n+1}^i, \dots, A_{p_i}^i)$  ( $p_i \geq n$ ) de karakteristieken van de entiteitverzamelingen  $\delta E_i$  ( $i = 1, \dots, m$ ) die alle de attributen  $\{A_1, \dots, A_n\}$  gemeenschappelijk hebben. Een entiteitverzameling  $\delta G$  is een *generalisatie* van  $\delta E_1, \dots, \delta E_m$  gedefinieerd als  $G(A_1, \dots, A_s, A_{s+1}, \dots, A_n, C)$ , indien er een *categorie-attribuut*  $C: \delta G \rightarrow \{E_1, \dots, E_m\}$  bestaat, dat elke entiteit uit  $\delta G$  afbeeldt op één en slechts één naam. Deze naam correspondeert met precies één van de onderliggende gegeneraliseerde deelverzamelingen. Het attribuut  $C$  heet ook wel het *onderliggende attribuut*. Omdat de onderliggende entiteitverzamelingen disjunkt moeten zijn, geldt:  $\delta G = \delta E_1 \cup \dots \cup \delta E_m$  en  $\delta E_i \cap \delta E_j = \emptyset \forall i \neq j$ . Een hiërarchie heet *compleet* als voor elke entiteit in de superverzameling een corresponderende entiteit bestaat in één van de direkt onderliggende entiteitverzamelingen. De pijl naar de superverzameling is dan donker ingekleurd.<sup>41</sup>

Een voorbeeld van generalisatie is te zien in figuur 11, waar PRODUKT (NAAM, SOORT) een gegeneraliseerd concept weergeeft, bestaande uit een ingekocht produkt (GRONDSTOF), een tussentijds produkt (HALFPRODUKT) een een eindprodukt (EINDPRODUKT), elk met eigen kenmerken. Te definiëren is derhalve de afbeelding

$$\text{SOORT: } \delta \text{PRODUKT} \rightarrow \{\text{GRONDSTOF, HALFPRODUKT, EINDPRODUKT}\}$$

en de volgende vier entiteitverzamelingen:

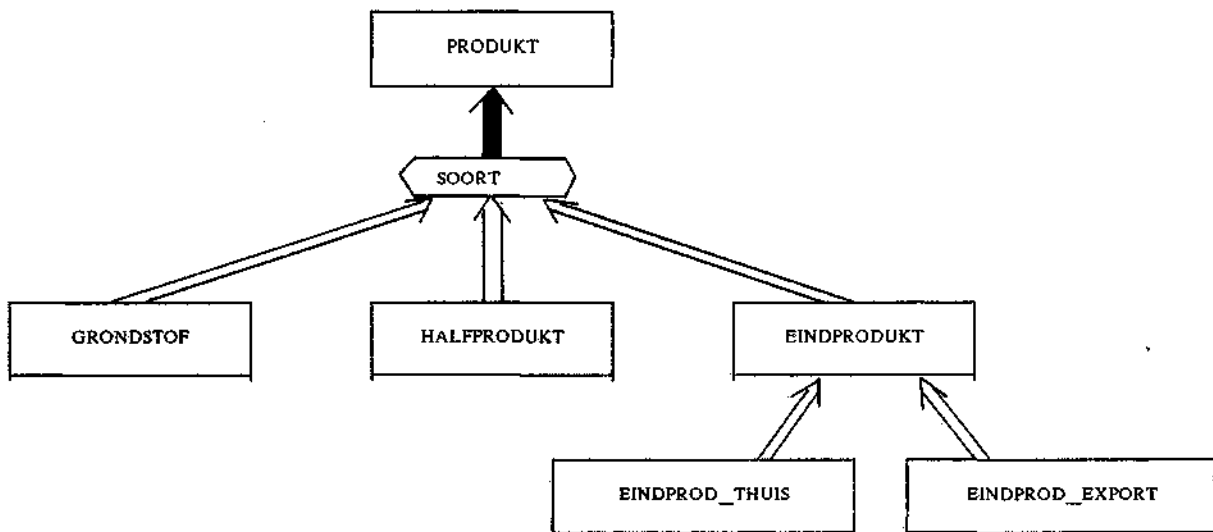
PRODUKT (NAAM, SOORT),

GRONDSTOF (NAAM, VERREKENPRIJS, LEVERANCIER, INKOOPTPRIJS).

HALFPRODUKT( NAAM , VERREKENPRIJS) en  
EINDPRODUKT ( NAAM , VERREKENPRIJS).

Op grond van het categorie-attribuut SOORT wordt elke entiteit uit PRODUKT afgebeeld naar één van de onderliggende deelverzamelingen. Bijvoorbeeld SOORT ("Etheliet") = GRONDSTOF en de entiteit "Etheliet" is bijgevolg ondergebracht in de deelverzameling  $\delta$ GRONDSTOF met z'n specifieke attributen. Het onderliggende attribuut vormt derhalve de verbinding tussen de deelverzamelingen en de bovenliggende entiteitverzameling en heeft een aparte plaats in het U-E-R diagram.

Er is overigens een opmerkelijke overkomst tussen het categorie-attribuut C en de functie TYPE, immers  $C = TYPE \mid \delta G$ ; een categorie-attribuut is gedefinieerd overeenkomstig de TYPE-functie, doch beperkt tot de elementen uit de verzameling  $\delta G$ .

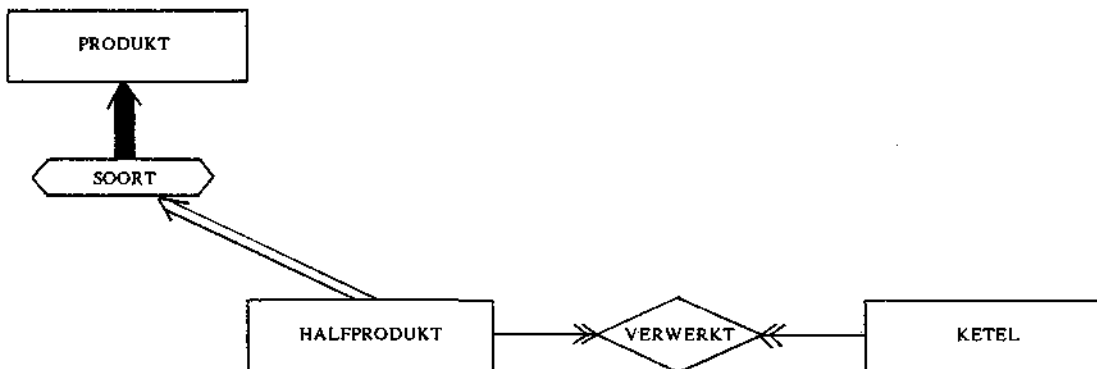


Figuur 11. Generalisatie hiërarchie en deelverzameling hiërarchie.

Tenslotte kan nog worden opgemerkt dat een bijzondere vorm van beperktheid (zie paragraaf 4.2) kan worden ondervangen door gebruik te maken van een hiërarchie. Het diagram in figuur 5.a is hiervan een voorbeeld. Enerzijds vanwege de binaire relatievorm, anderzijds vanwege het uniforme karakter van de restrictie, immers hier geldt:

$$\forall e_i, e_j \in \delta \text{KETEL} (i \neq j): C^{e_i} = C^{e_j} = \delta \text{HALFPRODUKT} \subset \delta \text{PRODUKT}$$

Door  $\delta \text{KETEL}$  te relateren aan  $\delta \text{HALFPRODUKT}$  waarvan  $\delta \text{PRODUKT}$  de superverzameling is, kan de beperktheid worden geëlimineerd (zie figuur 12).



Figuur 12. Eliminatie van beperktheid.

## 6. Samenvatting en conclusies

De basis van dit artikel wordt gevormd door het onderscheid tussen een entiteit (bijvoorbeeld een specifiek produkt), een verzameling entiteiten (PRODUKTEN) en het type van deze entiteiten (PRODUKT). Bovendien wordt een duidelijk onderscheid gemaakt tussen een verzameling entiteiten en de *karacterisering* van deze verzameling. Een entiteitverzameling wordt gekarakteriseerd door zijn type en zijn attribuut-waardenverzameling-paren. Associaties tussen entiteitverzamelingen worden, qua typering en karakterisering op eenzelfde wijze behandeld. Deze basis voor entiteiten, attributen en relaties werkt door in alle daaropvolgende definities en (mathematische) konstrukties.

De eerste uitbreiding, bestaansafhankelijkheid, wordt gezien als een eigenschap van de combinatie van entiteitverzamelingen in relatie tot elkaar. Het bestaan van een entiteit is vastgelegd door zijn 'existentie-deelverzameling'. De tweede uitbreiding betreft de deelname van een entiteit in een relatieverzameling. Hier wordt onderscheid gemaakt tussen het maximum aantal relaties waarin de entiteit mag participeren (kardinaliteits-klasse), de beperking van deelname aangegeven door de 'restriktie-deelverzameling' (restriktie-klasse), en de minimum deelname in de relatieverzameling (participatie-klasse).

Teneinde verschillende niveaus van detaillering tijdens het ontwerpproces toe te staan, zijn onder de noemer 'abstracties' een aantal structuurprimitieven ingevoerd. De grafische weergave van attributen kan in een eerste-fase ontwerp achterwege worden gelaten. Evenzo is de weergave van (complexe) deelstructuren op dit niveau uit te stellen door gebruik te maken van aggregatie en generalisatie. Een aggregaat is een relatieverzameling gezien als entiteitverzameling. In dit artikel wordt aangetoond dat aggregatie, zo gezien, een wiskundig te verantwoorden konstruktie is. Generalisatie tenslotte, wordt onderscheiden in deelverzameling hiërarchie en (exklusieve) generalisatie. Beide vormen zijn dankzij het eerder gelegde fundament eenvoudig te verantwoorden.

Een beslissende, weliswaar slecht te definiëren faktor in het maken van keuzen tussen alternatieve grafische representaties van primitieven in het hier opgebouwde Uitgebreide-Entiteit-Relatie model, is de bruikbaarheid van deze representaties.

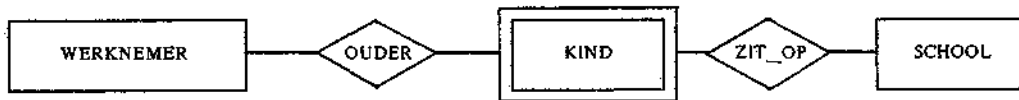


## 7. Referenties en notenlijst

- 1 Het hiërarchische model is beschreven door M.E. Senko & E.B. Altman & M.M. Astraham & P.L. Fehder: *Data structures and accessing in data-base systems*. In: IBM Systems Journal, Vol. 12, No. 1, 1973, blz. 30-93.  
Het netwerk model beschreven door C.W. Bachman: *Data structure diagrams*. In: Database, Vol. 1, No. 2, 1969, blz. 4-10.  
Het relationele model is van E.F. Codd: *A relational model of data for large shared data banks*. In: Communications ACM, Vol. 13, No. 6, 1970, blz. 377-387.
- 2 [Chen76].
- 3 Voor de verscheidene uitbreidingen van het oorspronkelijke E-R model vinden we in de literatuur vele benamingen als Extended-E-R (EER of XER) model bij onder andere [Scheu], [Schif] en ook [NavCh], hoewel allen enkele ietwat afwijkende definities hebben. Ook het Enriched-E-R (EER) model komen we regelmatig tegen in onder andere [Dataid1], [Bati84], [Busso] en [LenSan]. Daarnaast bestaan er nog vele andere afwijkende vormen als Enhanced-Entity-Relation, Entity-Attribute-Relation (EAR) en Entity-Category-Relation (E-C-R) model.
- 4 [VUMAS].
- 5 De definitie van een entiteit (in [Chen76; blz.10]) wordt letterlijk geciteerd door [Schif; blz.140] en [Sakai; blz.216]. In [LenSan; blz.531] representeert een entiteit een verzameling objekten ("instances") met dezelfde eigenschappen. Doch dit is geen gangbaar taalgebruik.
- 6 [Chen76; blz.11] spreekt over een predikaat dat is geassocieerd met elke entiteitverzameling om te testen of een entiteit ertoe behoort. Dit predikaat komt in feite overeen met de naam van de verzameling en hij rekent het tot één van de eigenschappen van een entiteitverzameling.
- 7 De begrippen entiteit, entiteitverzameling (en entiteittype) worden doorgaans nogal verschillend gebruikt. Wij handhaven de terminologie zoals door [Chen76] geïntroduceerd en door [Sakai] verder geformaliseerd:  
*"Entities can be classified into different entitytypes ...  
An entityset is a group of entities of the same type."*  
[Chen77; blz.78-79] verdedigt dat de begrippen "verzameling" en "type" door elkaar mogen worden gebruikt.
- 8 Attributen behoren tot de derde categorie van basisbegrippen van [Chen76; blz.12], overgenomen door [Sakai; blz.216]. Een attribuut wordt in [LenSan; blz.531] gedefinieerd als "een korrespondentie tussen entiteiten (of relaties) en waardenverzamelingen (het domein van het attribuut) die de maat van de eigenschap representeren."  
Gebruik makende van de TYPE-functie, gedefinieerd op bijbehorende klassen, zou hier kunnen worden volstaan met een generaliserende definitie van een attribuut A als functie van ofwel een entiteitverzameling  $\delta N$  ofwel een relatieverzameling van het type N naar een waardenverzameling (of cartesisch produkt van waardenverzamelingen) V, gedefinieerd als  $A: \delta N \rightarrow V$ .  
Omdat in het U-E-R diagram een attribuut anders van vorm wordt dan in het E-R diagram van [Chen76], bespreken we hier nog niet de weergave ervan. [Webre; blz.190] vindt de manier van [Chen76] zelfs veel te complex.
- 9 [Sakai; blz.216].
- 10 In [Chen76; blz.16] en [Sakai; blz.216] wordt gesproken over een *entiteitupel* uit een tabel die een *entiteit-relatie* wordt genoemd.
- 11 Identificerende attributen en sleutels worden genoemd in [Chen76; blz.14] en de meer formele beschrijving ervan in [Sakai; blz.216] en [LenSan; blz.531], hoewel bij de laatste de term "identifler" wordt gebruikt en men net als [Dataid1; blz.9] met de termen "internal" en "external identifler" reeds vooruit loopt op bestaansafhankelijkheid.
- 12 Opmerkingen over het begrip relatie in het algemeen maken o.a. [Chen76; blz.10], [Dave; blz.607], [LenSan; blz.531] en [Schif; blz.140].
- 13 Relatieverzameling en relatietype worden meer formeel omschreven in [Chen77; blz.79], [Sakai; blz.216-217] en [Schif; blz.140]. Daarnaast toont [Chen76; blz.17-18] een tabel, "relationship

relation" genaamd, waarbij elke rij van waarden een "relationship tuple" heet. De term "relationship" wordt in de literatuur gebruikt voor zowel een relatie als voor een relatieverzameling (of -type). Wij beschouwen daarentegen een relatie  $r$  als een element van een relatieverzameling  $\delta R$ . [Sakai; blz.217] noteert een relatie  $r$  als  $r = (e_1, \dots, e_m, v_1, \dots, v_n)$  met  $e_i \in \delta E_i$ ,  $v_j \in V_j$  en  $v_j = A_j(e_1, \dots, e_m)$  ( $i = 1, 2, \dots, m$ ;  $j = 1, 2, \dots, n$ ) en geeft bovendien zijn schrijfwijze voor een relatieverzameling en de primaire sleutel ervan. Evenals bij entiteitklassen zullen in dit artikel relatieklassen worden beschouwd als relatieverzamelingen en genoteerd als  $\delta R$ .

- 14 Een relatieverzameling  $R$  is, aldus [Chen76; blz.11-12],  
"... a mathematical relation among  $m$  entities, each taken from an entity set:  
 $\{[e_1, \dots, e_m] \mid e_1 \in E_1, \dots, e_m \in E_m\}$ , and each tuple of entities,  $[e_1, \dots, e_m]$  is a relationship."
- 15 Deze formele beschrijving is van [Sakai; blz.216] aan de hand van de tekst uit [Chen76; blz.12].
- 16 Dit behoort bij het oorspronkelijke E-R model van [Chen76; blz.18,20 e.v.] en wordt door velen (waaronder [Chen77], [Webre], [Schif] en [Scheu]) gevolgd. De aanduiding wordt ook wel "mapping class" genoemd.
- 17 In [NavCh; blz.223-224] wordt verwezen naar recent onderzoek naar semantische datamodellen. Men vindt het E-R model in de huidige vorm nog onvoldoende vanuit semantisch standpunt (en ze breiden het daarom ook uit).  
[Schif; blz.217] en [Scheu; blz.121-122] noemen de gelijkenis van het E-R model met de semantiek van natuurlijke taal als een belangrijk voordeel, maar het gemis van uitdrukkingsmogelijkheden van abstracties en meerdere visies als een nadeel.  
[Webre; blz.173] zag meer gebreken op het gebied van klassifikatie van relatietypen en de afhankelijkheid van entiteiten.
- 18 [Chen76; blz.18,20] kent al "existency dependency" maar hij combineert dit tegelijkertijd met "weak entity relations" (waarbij een relatie wordt gebruikt om entiteiten te identificeren). Zijn definitie van bestaansafhankelijkheid komt overeen met die van [Webre; blz.175-176] en slaat voornamelijk op de semantische betekenis van afhankelijkheid van entiteiten. De tweede vorm van bestaansafhankelijkheid volgens [Chen76; blz.17] wijkt -- zoals ook in [Scheu; blz.125] wordt vermeld -- iets af van de oorspronkelijke definitie en is gebaseerd op de situatie dat attributen zowel entiteiten van de eigen entiteitverzameling als van een andere -- de afhankelijke of zwakke -- entiteitverzameling identificeren.  
Omdat de uiteindelijke definitie van bestaansafhankelijkheid volgens [Webre; blz.176] toepasbaar is op beide gevallen, verdient die de voorkeur.  
Zie ook noot 20.
- 19 Hier is gekozen voor de notatie van [Chen76; blz.20] en [Webre; blz.179,186,193], omdat deze de twee vormen van bestaansafhankelijkheid (zie noot 18) dekt. Kijkend naar de notatie van een zwakke entiteitverzameling volgens [Chen76; blz.17,20], stuiten we op een probleem dat is uitgebeeld in figuur 13, waarin niet meer is te herkennen welke de zwakke en welke de sterke entiteitverzameling voorstelt (zie ook [Webre; blz.175]). Ook de notatie van [Scheu; blz.126] is afwijkend en past niet op alle situaties. Bovendien is er een verschil van mening of bestaansafhankelijkheid kenmerkend is voor de entiteitverzameling ([Webre; blz.175-176]), of voor de relatieverzameling ([Scheu; blz.125] en [Chen76; blz.17,20]). Dit houdt uiteindelijk verband met de twee vormen van bestaansafhankelijkheid.  
Onze mening is dat bestaansafhankelijkheid een eigenschap is die niet alléén aan de entiteitverzameling of alleen aan de relatieverzameling kan worden toegeschreven maar meer berust op een combinatie van beide. De voor dit artikel gekozen notatie van [Chen76; blz.20] en [Webre; blz.197,186,193] past hier dan ook het beste bij. De laatste auteur gebruikt voor de bestaansafhankelijkheden een "existence digraph", een apart schema ter bestrijding van de toenemende complexiteit van het conceptuele schema.



Figuur 13.

20 We citeren [Busso; blz.28]:

*"An internal identifier is an attribute or group of attributes that uniquely determines an entity; entities may also be identified through other entities associated with them (external identifier). An entity may have many identifiers (internal and/or external)."*

Ook [LenSan; blz.533-534] geeft een definitie van een "identifier" die is uitgebreid tot externe identificatie. De representatie hiervan (zie paragraaf 5.1) is te vinden in onder andere [Busso], [Bati83] en [Bati84].

21 Voor [Webre; blz.175-176] was dit een van de redenen voor de herdefinitie van bestaansafhankelijkheid.

22 De naamgeving van klassen wijkt af van [Webre; blz.177-181,193]. "Mapping class" suggereert dat een relatieverzameling een afbeelding is. De term "kardinaliteits-klasse" is een betere aanduiding van de maximum deelname in een relatieverzameling. Ook in [LenSan; blz.531] wordt afwijkend gedefinieerd dat

*"Cardinality constraints are restrictions on the minimum and the maximum number of instances of a class (or a set of classes) that are related to the instances of another class (or set of classes)."*

Hiermee combineert hij twee verschillende semantische beperkingen van een relatietype. De definitie van "Relationship completeness class" van [Webre; blz.178-181] omvat de minimum deelname restrictie van [LenSan; blz.531] en onderscheidt bovendien partiële, totale en complete deelname. Deze uitbreiding leidt tot onderkenning van deze drie hoofdklassen.

23 Motief hiervoor is het gebruik van grafische pakketten waarmee een vorm van het U-E-R diagram op een beeldscherm kan worden verwerkt. De notatie van "1" of "M" zal waarschijnlijk praktische problemen opleveren en is ook niet erg overzichtelijk.

Er zijn vele notatievormen mogelijk. We verwijzen onder andere naar [Busso] en [Bati83] die pijltjes op de verbindingslijnen gebruiken en naar [Dataid1; blz.17] waarin dit ook gebeurt, maar waar de pijltjes precies andersom staan genoteerd. Geheel anders werken [Sakai] en [Bati84] met "max-card" aanduidingen. Als max-card = 1 wijst dit volgens [LenSan; blz.540] op een functionele afhankelijkheid tussen twee entiteit(deel)verzamelingen. Omdat de "pijltjes-notatie" van [Dataid1] efficiënt en ons inziens ook elegant is, kiezen wij deze vorm voor het U-E-R diagram.

24 Restriktie-klasse is te vinden in [Webre; blz.178-180,193].

25 Met een analoge inleiding introduceert [Webre; blz.179-180] de situatie "existence restricted".

26 In eerste instantie worden in [Scheu; blz.124-125] de begrippen "totaal" en "partieel" geïntroduceerd. Het is overgenomen door onder andere [Dataid1; blz.9] en [LenSan; blz.540-541] ofschoon [Webre; blz.177-181] er uiteindelijk een meer samenhangend geheel van heeft gemaakt. De "mapping ratio" van [Sakai] en de analoge notatie van [Bati84] bevatten daarentegen reeds de gegevens over participatie-klasse (zie ook de derde alinea van noot 27).

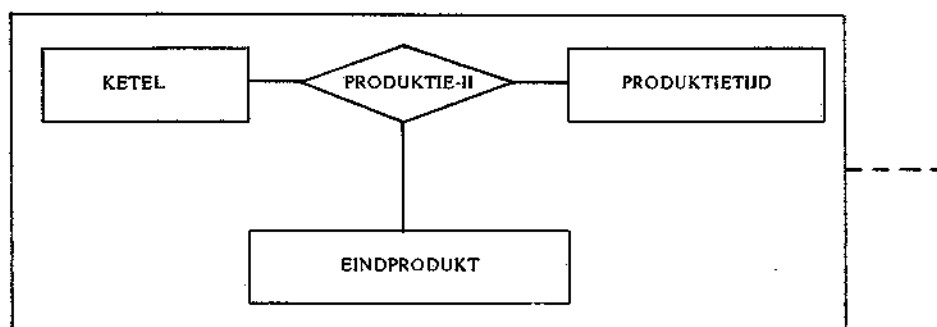
27 [Webre; blz.178-181] is namelijk de enige die een relatieverzameling kent die "compleet" kan zijn. Dit heeft te maken met de restrictieverzameling waarop hij de participatie-klasse definieert wat bij andere auteurs niet wordt gedaan. In de overige literatuur vinden we dan ook slechts de vormen totaal en partieel waarvoor echter uiteenlopende notatievormen zijn gekozen.

Allereerst zijn er de aanduidingen van [Scheu; blz.124-126 e.v.] en [Webre; blz.180-181,185,193]. Deze voldoen weliswaar, maar kunnen niet worden gekombineerd met de grafische kenmerken van de kardinaliteits-klasse.

Een andere veel gebruikte vorm is die van "min-card(inalities)" en "max-card(inalities)" volgens onder andere [Sakai] en [Bati84] en op vergelijkende wijze [LenSan]. Max-card komt overeen met de kardinaliteits-klasse, het maximum aantal keer dat elke entiteit in relaties mag voorkomen. Min-card is daarentegen het minimum aantal dat elke entiteit in een relatie moet zijn betrokken. Een

min-card van 1 komt overeen met de situatie "totaal" in de participatie-klasse. Hoewel op deze wijze de betreffende semantische informatie over de relatieverzameling goed wordt weergegeven, is een bezwaar tegen de notatie-vorm op zijn plaats. De aanduidingen worden los aan het relatietype-symbool toegevoegd, hetgeen de duidelijkheid niet ten goede komt. Voorkeur verdient een eenvoudige toevoeging aan de relatieverzameling *alleen* in het geval deze partieel is: een dwarsliggend streepje op de verbindingslijn aan de kant van de entiteitverzameling ten opzichte waarvan de relatieverzameling partieel is. Met deze representatie voor de participatie-klasse, die overeenkomt met de notatie van attributen (zie paragraaf 5.1) volgen wij [Busso] en [Bati83].

- 28 [Webre; blz.179-180].
- 29 [SmAgg] en [SmGen] introduceerden het begrip abstractie zoals hier toegepast. Behalve het onderdrukken van details komt vanzelfsprekend ook aan de orde wanneer het vermelden van details in het U-E-R diagram wel van belang is.
- 30 Aggregatie naar een idee van [SmAgg; blz.405-406]. Zie ook [Scheu; blz.123], [Webre; blz.185-189] en [Schif; blz.144].
- 31 De hier bedoelde hiërarchieën vinden hun oorsprong in [SmGen].
- 32 Zie [Schif; blz.140]. [Webre; blz.190] gaat zelfs nog verder en wil in het geheel geen weergave van attributen in het U-E-R diagram. Dit wordt te complex en hij ziet daarom meer heil in de notatie van deze gegevens op aparte lijsten.
- 33 Deze notatie, die nogal afwijkt van de oorspronkelijke van [Chen76; blz.14-16] wordt door velen gebruikt: [Schif; blz.140], [Scheu; blz.125-126], [LenSan], [Busso] en [Bati84].
- 34 "Repeating attributes" of "multivalued attributes" zijn te vinden in [Busso; blz.28-29,43], [Bati83; blz.76] en [Dataid1; blz.9,17]. De analogie met de pijltjes van kardinaliteits-aanduidingen spreekt voor zich.
- 35 De toepassing van de "optionalities definition" uit [Busso; blz.33] wordt ook gebruikt in onder meer [Bati83] en [Bati84]. De achterliggende analogie wordt in [Dataid1; blz.9] verwoord:  
*"Both attributes and relationships can be partial (optional) or total."*
- 36 De hier gebruikte notatie is te vinden in [Webre; blz.185 figuur 4.1 en blz.193 figuur A-1]. Een rechthoek ter grootte van een entiteitverzameling heeft meer grafische expressiekracht en ligt meer in de lijn van de definitie dan het alles omvattende blok van [Scheu; blz.123-124,136], [Schif; blz.144] en [NavCh; blz.234,239,241]. Niet zozeer de entiteitverzameling(en) als wel de relatieverzameling wordt namelijk geaggregeerd. Het verschil is te zien in figuur 14 (vergelijk figuur 9). Bovendien lijkt het alsof de entiteitverzamelingen binnen het blok niet meer mogen worden gerelateerd met erbuiten liggende concepten.



Figuur 14. Een alternatieve notatie voor een geaggregeerde relatieverzameling  $\delta$ PRODUKTIE-II.

- 37 Met deze notatie volgen we [LenSan; blz.531], [Busso; blz.28] en [Bati84; blz.652] en [Dataid1; blz.18].
- 38 Deze definitie is van [NavCh; blz.228]. In [LenSan; blz.531], [Busso; blz.28] en [Bati84; blz.652] en [Dataid1; blz.18] wordt iets vergelijkbaars geformuleerd, namelijk "F is een deelverzameling van

een andere entiteitverzameling  $E$  (de superverzameling) als elke entiteit  $f$  uit  $F$  ook in  $E$  zit ( $\forall f \in F: f \in E$ ).

In [LenSan; blz.532-533] wordt voorgesteld om het gebruik van deelverzameling hiërarchieën uit te breiden teneinde deze ook toe te passen op de relatieverzamelingen. Met behulp van het in dit artikel gedefinieerde begrip aggregatie is dat overbodig.

- 39 De basis van generalisatie is gelegd door [SmGen]. Verbeterde versies, meer passend in de context van het U-E-R model zijn van [Schif; blz.143] maar vooral van [LenSan; blz.531], [Busso; blz.28-43] en [Bati84; blz.652-663], [NavCh; blz.230,233,239] en [Dataid1; blz.18].
- 40 De deelverzameling hiërarchie is te vergelijken met de uitdrukking "*ISA\_hierarchy*" ([NavCh; blz.232]).
- 41 Zie [Busso; blz.28-29].

## 8. Literatuur

- [Bati83] C. Batini & M. Lenzerini & M. Moscarini, *Views integration*. In: *Methodology and Tools for Data Base Design*, S. Ceri (ed.), 1983, blz. 57-84.
- [Bati84] C. Batini & M. Lenzerini, *A Methodology for Data Schema Integration in the Entity Relationship Model*. In: *IEEE Transactions on Software Engineering*, Vol. SE-10, No. 6, november 1984, blz. 650-664.
- [Busso] U. Bussolati & S. Ceri & V. De Antonellis & B. Zonta, *Views conceptual design*. In: *Methodology and Tools for Data Base Design*, S. Ceri (ed.), 1983, blz. 25-55.
- [Chen76] P.P-S. Chen, *The Entity-Relationship Model - Toward a Unified View of Data*. In: *ACM Transactions on Database Systems*, Vol. 1, No. 1, maart 1976, blz. 9-36.
- [Chen77] P.P-S. Chen, *The Entity-Relationship Model - A basis for the enterprise view of data*. In: *AFIPS*, Vol. 46, 1977, blz. 77-84.
- [Dataid1] V. De Antonellis & A. Di Leva, *DATAID-1: A Database Design Methodology*. 1984.
- [Dave] R.A. Davenport, *The Application of Data Analysis - Experience with the Entity-Relationship Approach*. In: *Entity-Relationship Approach to Systems Analysis and Design*, P.P. Chen (ed.), 1980, blz. 603-621.
- [LenSan] M. Lenzerini & G. Santucci, *Cardinality Constraints in the Entity-Relationship Model*. In: *Entity-Relationship Approach to Software Engineering*, C.G. Davis & S. Jajodia & P.A. Ng & R.T. Yeh (eds.), 1983, blz. 529-549.
- [NavCh] S.B. Navathe & A. Cheng, *A Methodology for Database Schema Mapping from Extended Entity Relationship Models into the Hierarchical Model*. In: *Entity-Relationship Approach to Software Engineering*, C.G. Davis & S. Jajodia & P.A. Ng & R.T. Yeh (eds.), 1983, blz. 223-248.
- [Sakai] H. Sakai, *On the optimalization of an Entity-Relationship Model*. In: *AFIPS Conference Proceedings, 3rd USA-Japan Computer Conference*, 1978, blz. 218-220.
- [Schif] G. Schiffner & P. Scheuermann, *Multiple Views and Abstractions with an Extended-Entity-Relationship Model*. In: *Computer Languages*, Vol. 4, 1979, blz. 139-154.
- [Scheu] P. Scheuermann & G. Schiffner & H. Weber, *Abstraction Capabilities and Invariant Properties Modelling within the Entity-Relationship Approach*. In: *Entity-Relationship Approach to Systems Analysis and Design*, P.P. Chen (ed.), 1980, blz. 121-140.
- [SmAgg] J.M. Smith & D.C.P. Smith, *Database Abstractions: Aggregation*. In: *Communications of the ACM*, Vol. 20, No. 6, juni 1977, blz. 405-413.
- [SmGen] J.M. Smith & D.C.P. Smith, *Database Abstractions: Aggregation & Generalization*. In: *ACM Transactions on Database Systems*, Vol. 2, No. 2, juni 1977, blz. 105-133.
- [VUMAS] Faculteit der Economische Wetenschappen, *Vrije Universiteit Management Simulation*. Syllabus, oktober 1985.
- [Webre] N.W. Webre, *An Extended Entity-Relationship Model and its Use on a Defense Project*. In: *Entity-Relationship Approach to Information modelling and Analysis*, P.P. Chen (ed.), 1983, blz. 173-193.