



INF1300

Introduksjon til databaser

Dagens tema:

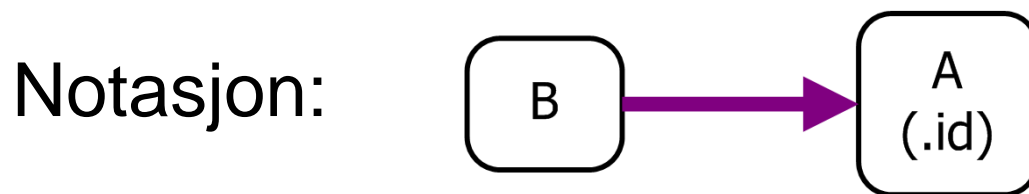
- **Underbegreper og underbegrepsskranker**
- **Kombinerte totale roller**
- **Ekvivalente stier og joinskranker**
- **Behandling av tid**
- **Informasjonsbærende representasjoner**
- **Undertrykking av begreper under gruppering**

Underbegreper

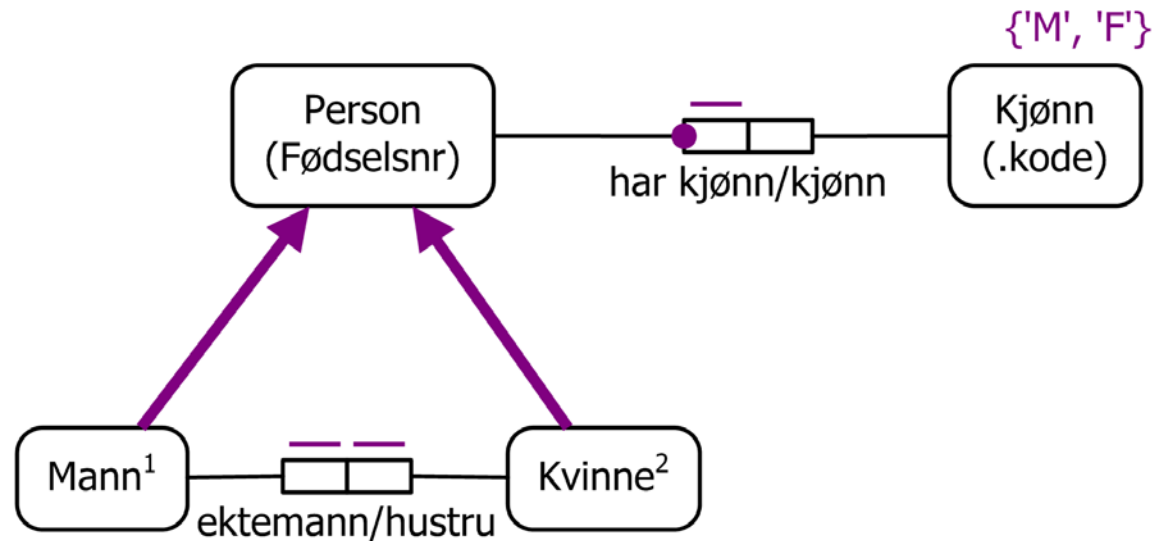
Kjernespørsmål: Kan alle tenkelige forekomster av et begrep spille alle roller som er knyttet til begrepet?

Hvis nei: Kan få en mer presis modell ved å innføre **underbegreper**

B er et underbegrep av A hvis vi alltid har at $\text{pop}(B) \subseteq \text{pop}(A)$



Underbegrepsskranke



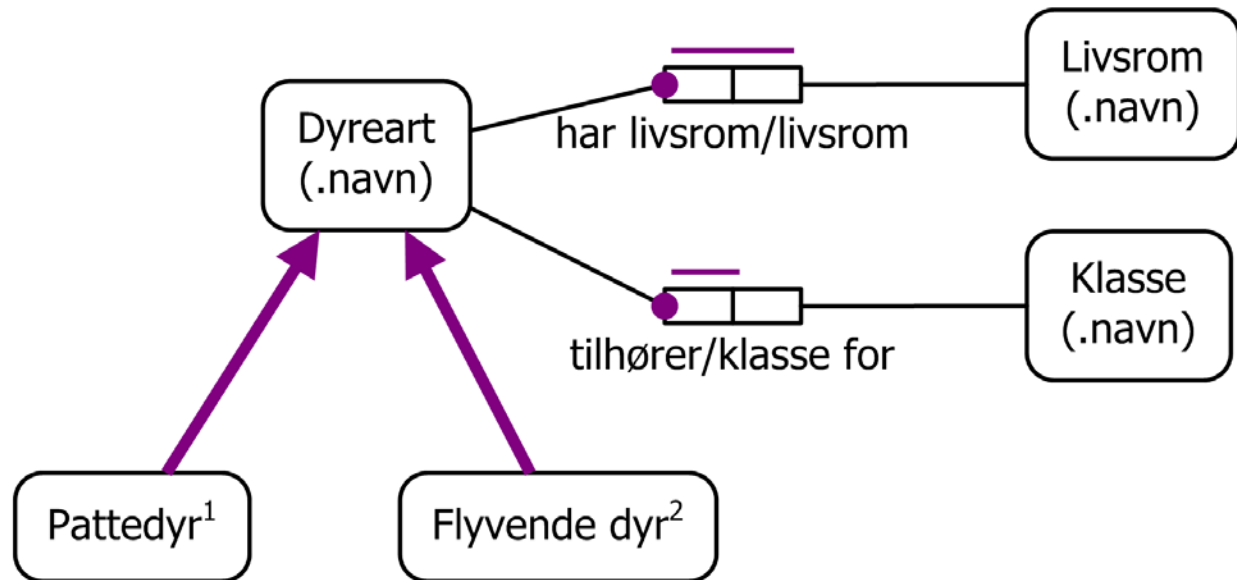
¹hver Mann er en Person som *har kjønn* 'M'

²hver Kvinne er en Person som *har kjønn* 'F'

Underbegreper

- **Underbegreper arver representasjon og roller fra superbegrepet.** I tillegg har de sine egne roller
- **Underbegrepsskranker** brukes til å bestemme hvilket underbegrep hver enkelt forekomst tilhører
- Underbegreper kan **overlappe** eller være **disjunkte**
- Underbegrepene *kan*, men *må* ikke, være **uttømmende** mhp. sitt superbegrep
- Resonnementer over entydighetsskranker, totale roller og underbegrepsskrankene avslører om underbegrepene er overlappende og/eller uttømmende

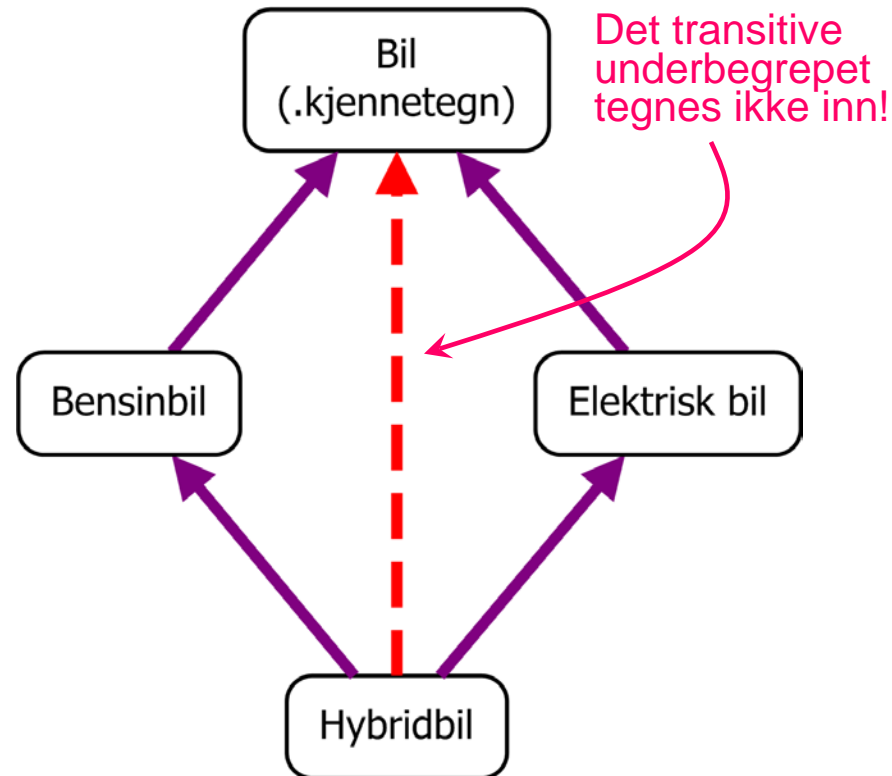
Eksempel på overlappende og ikke-uttømmende underbegreper



¹hvert Pattedyr er en Dyreart som *tilhører* Klasse 'Pattedyr'

²hvert Flyvende dyr er en Dyreart som *har livsrom* Livsrom 'Luft'

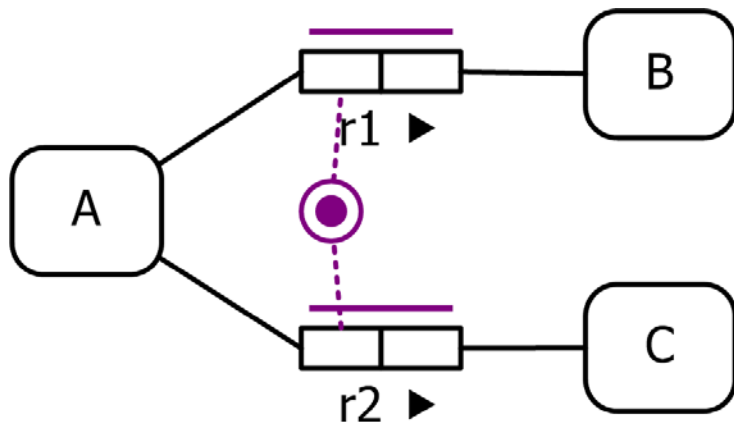
Underbegreper i flere nivåer



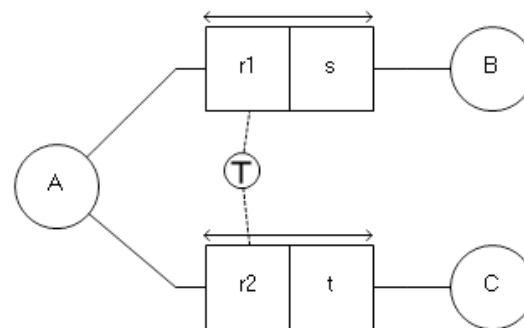
Kombinert total rolle

A skal ha enten rollen r1 eller rollen r2.

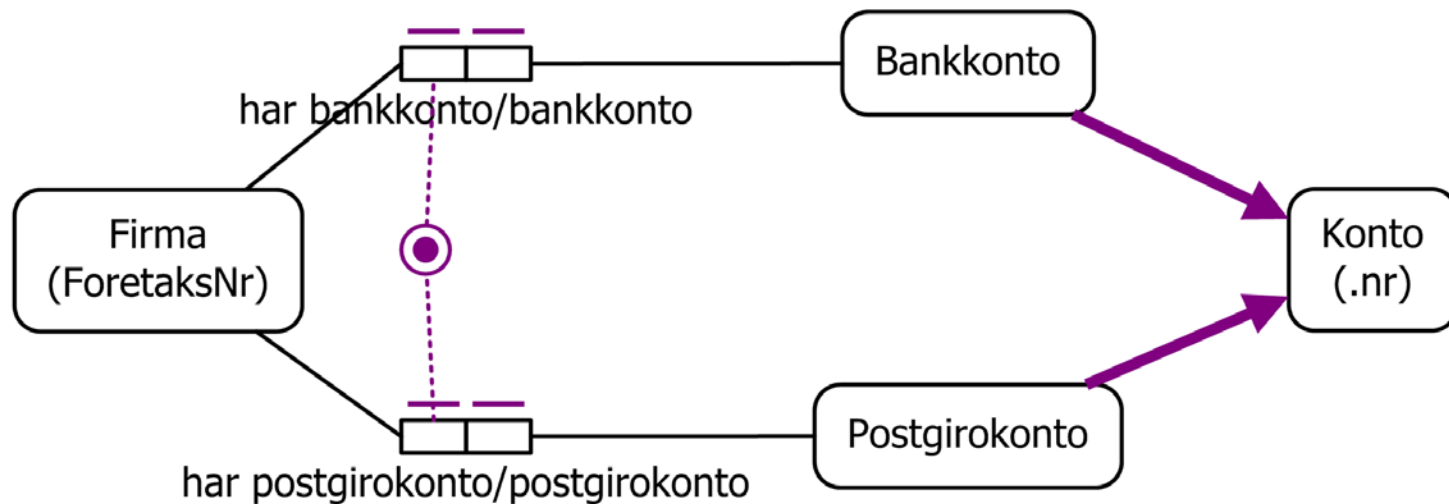
$\text{pop}(r1) \cup \text{pop}(r2) = \text{pop}(A)$ for alle tilstander



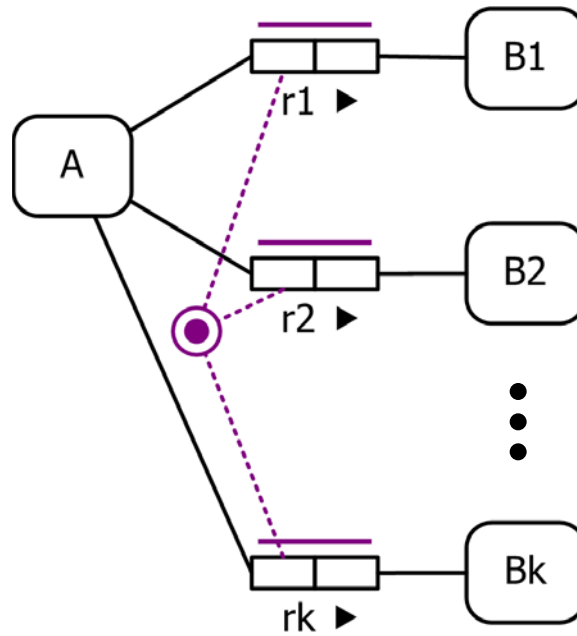
Kombinert total rolle i stORM:



Eksempel på kombineret total rolle og underbegrep



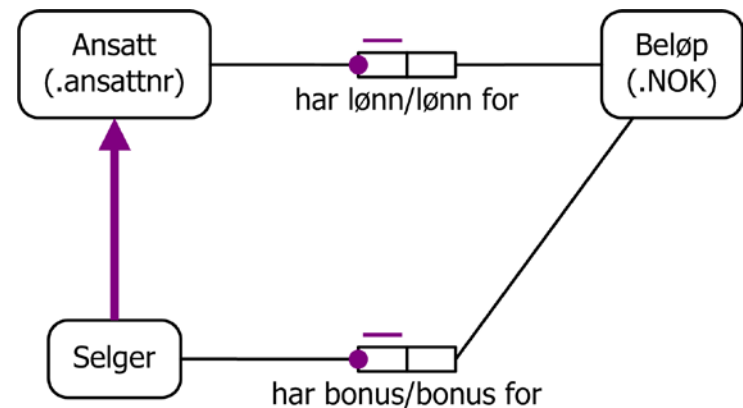
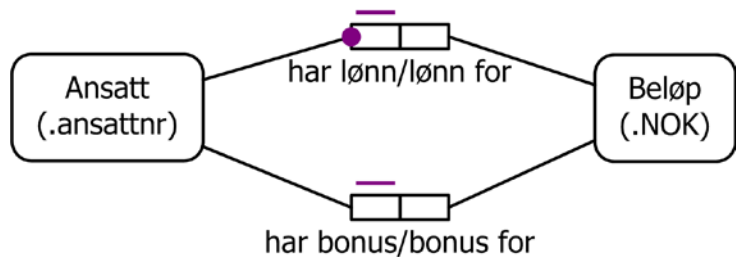
Generell kombiniert total rolle



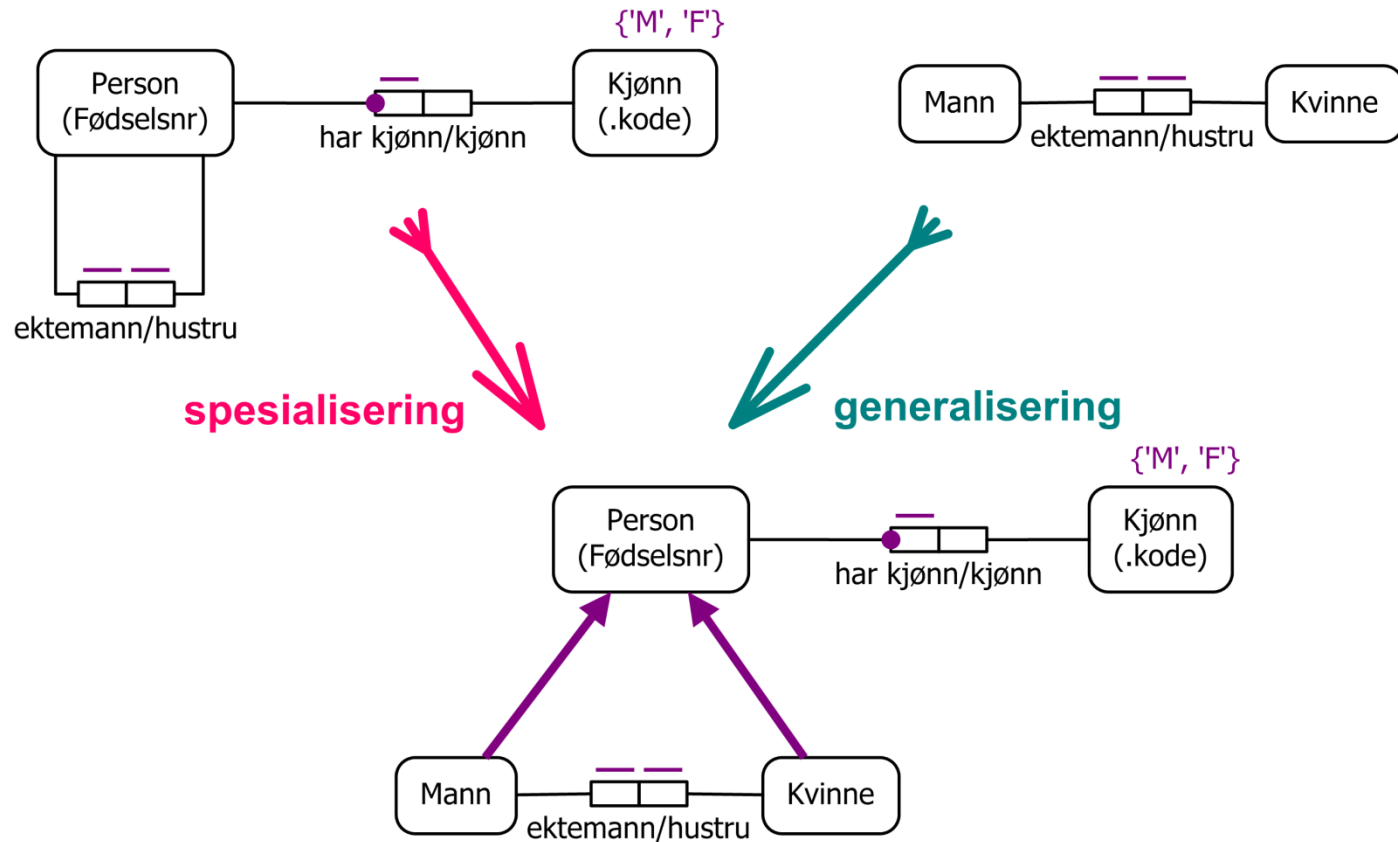
$$\text{pop}(r1) \cup \text{pop}(r2) \cup \dots \cup \text{pop}(rk) = \text{pop}(A)$$

Se opp for manglende totale roller

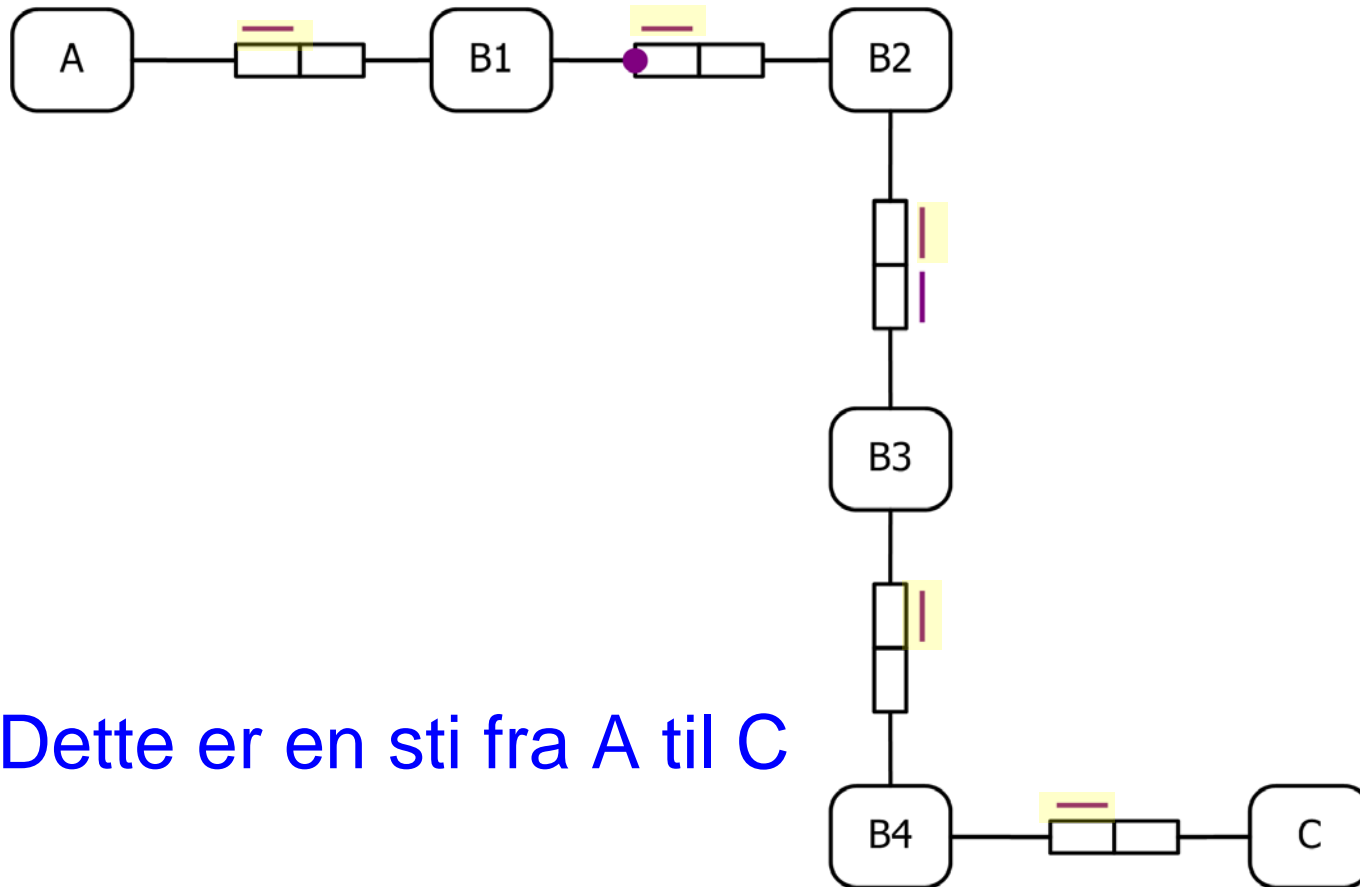
Mangel på totale roller kan indikere et underbegrep



Spesialisering og generalisering

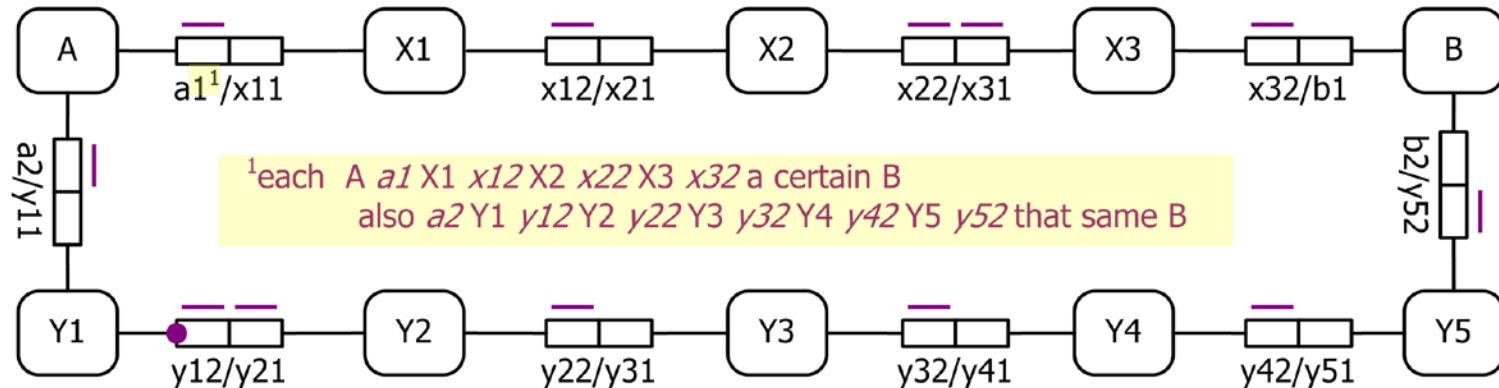


Stier



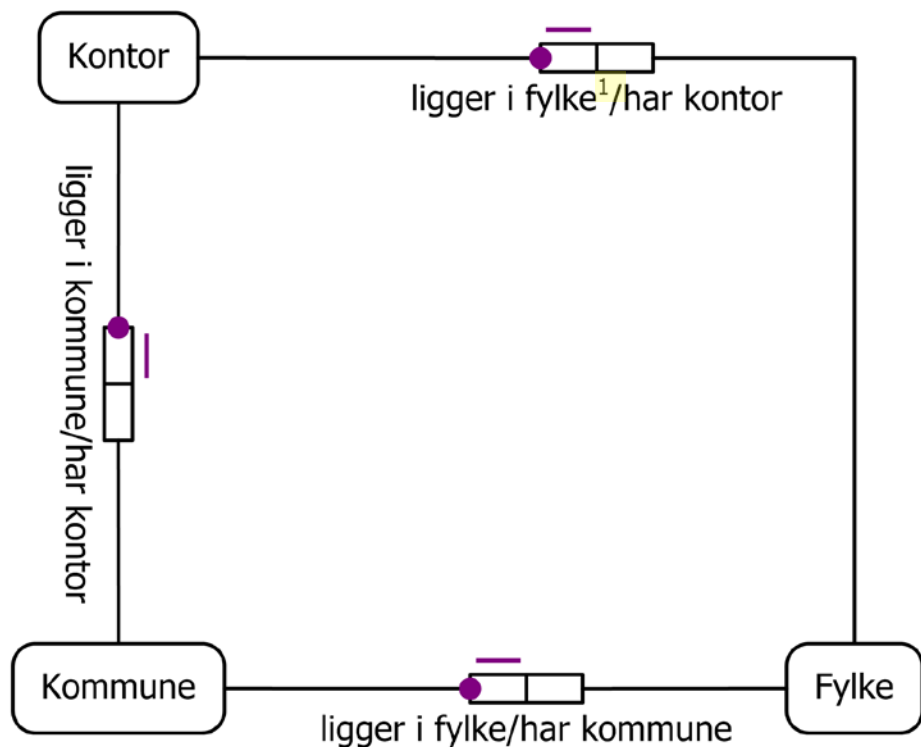
Dette er en sti fra A til C

Ekvivalente stier



Dersom vi har to stier fra A til B som er slik at hvis vi starter med én forekomst i A, så skal vi komme til *samme* forekomst i B uavhengig av hvilken av de to stiene vi følger, så kaller vi de to stiene **ekvivalente**.

Avledbare data - 1

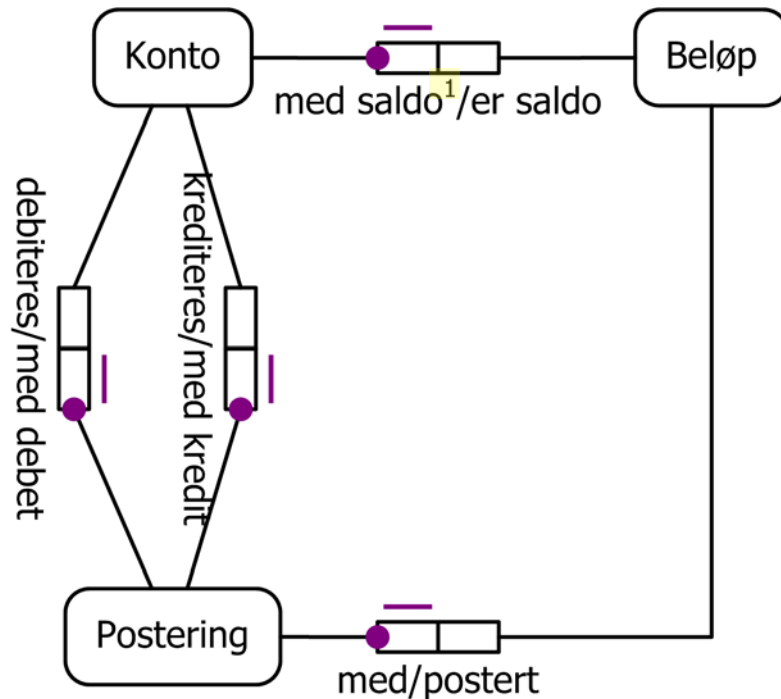


Unødvendig
(avledbar) informasjon

Merk de ekvivalente
stiene!

¹et Kontor som *ligger i fylke* et bestemt Fylke,
ligger i kommune Kommune som *ligger i fylke* det samme Fylket

Avledbare data - 2

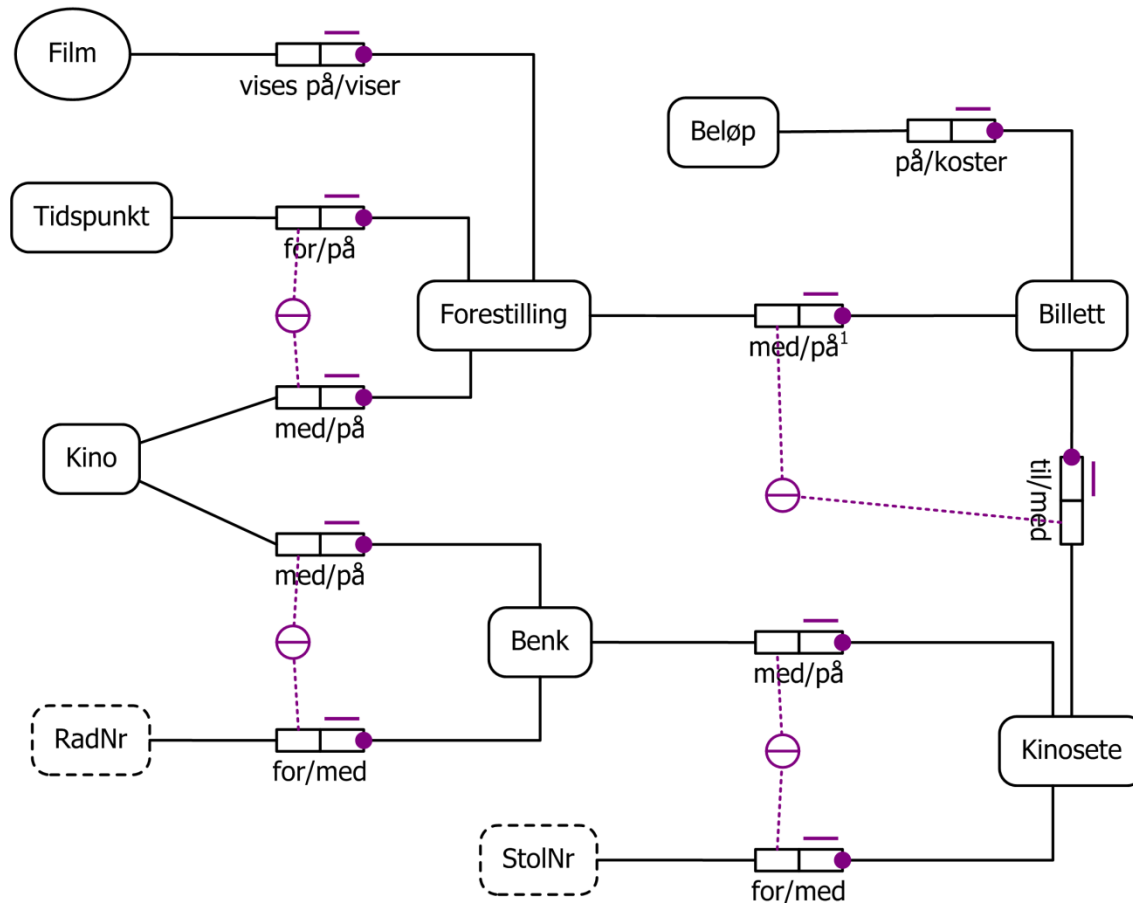


Unødvendig (avledbar) informasjon?

Det kan være mange grunner til å lagre avledbar informasjon

¹for hver Konto *er saldo* =
 $\text{sum}\{x \mid \text{Konto krediteres Postering med Beløp } x\} -$
 $\text{sum}\{y \mid \text{Konto debiteres Postering med Beløp } y\}$

Eksempel på ekvivalente stier



¹en Billett på en Forestilling på en bestemt Kino er til et Kinosete på en Benk på den samme Kinoen

Joinskranker

- Ekvivalente stier er et viktig eksempel på det vi kaller **joinskranker**
- For å håndheve disse skrankene kan det være nødvendig å foreta en join mellom tabeller
- Alle mengdeskranker kan opptre som joinskranker
- Ekvivalente stier er et spesialtilfelle av en joinlikhetsskranke

Avanserte skranker

- Alle skranker vi ikke har grafiske symboler for, kalles avanserte skranker
- Disse skrives på ORM-diagrammet som tekst
- Lovlig språk er førsteordens logikk og vanlig aritmetikk (regneformler)
- Lovlige variable er konstanter, roller, stier og inverse stier
- Se f.eks. lysarket Avledbare data – 2
- ORM 2 har et eget språk for avanserte skranker

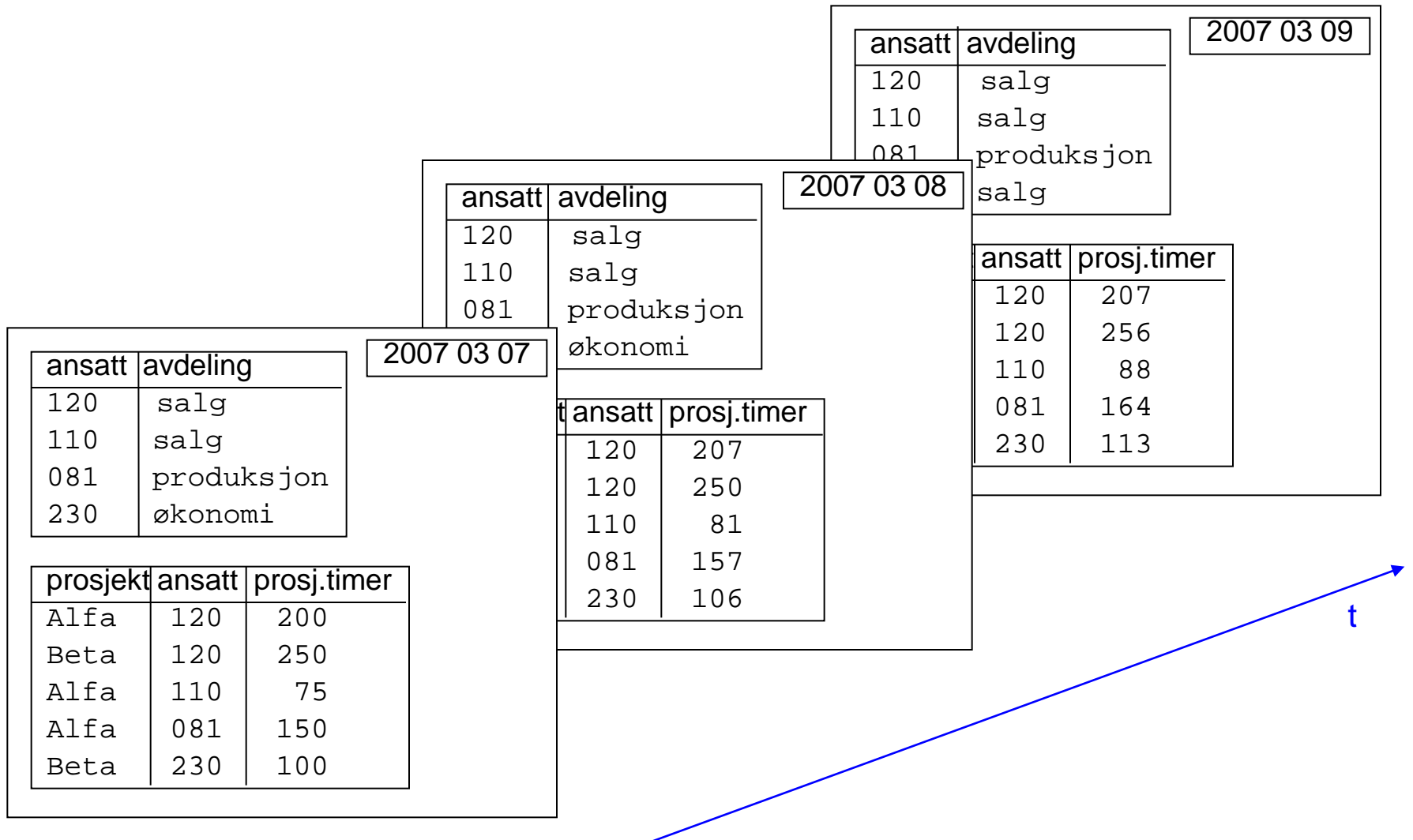
Behandling av tid

- Versjonering
- Hva er et tidspunkt
- Hva skal et tidsstempel reflektere
- Praktisk versjonsmodellering

Versjonering

- Hvis vi ønsker at databasen skal vise historiske opplysninger, lagrer vi **tidsstemplede versjoner** av informasjonen
- Med en **versjon** mener vi her et øyeblikksbilde av all informasjon
- De tidsstemplede versjonene kan ordnes langs en **tidsakse**

Versjoner med tidsstempel



Hva er et tidspunkt

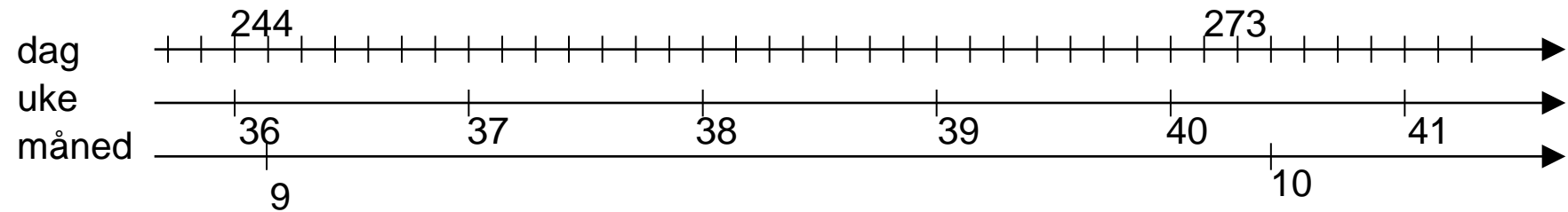
- Tidsaksen består i praksis **alltid** av **tidsintervaller** i informasjonsmodellen
- **Granulariteten** til intervallene avhenger av behovet for nøyaktighet. Granularitet avgjør
 - hvordan tidsintervallene skal representeres
 - Hvert tidsintervall identifiseres ved et **tidsstempel**, f.eks. år+måned, år+ukenummer, år+mnd+dag, år+mnd+dag+time+minutt
 - hva som er «samtidig»
 - Hendelser innen samme tidsintervall kan ikke skilles i tid
- En informasjonsmodell kan ha flere tidsakser med ulik oppdeling og granularitet
 - det er ikke alltid mulig å bestemme samtidighet på tvers av ulike tidsakser

Forskjellige tidsakser, eksempel

september 2009

| Uke | ma | ti | on | to | fr | lø | sø |
|-----|----|----|----|----|----|----|----|
| 36 | | 1 | 2 | 3 | 4 | 5 | 6 |
| 37 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
| 38 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
| 39 | 21 | 22 | 23 | 24 | 25 | 26 | 27 |
| 40 | 28 | 29 | 30 | | | | |

4: ○ 12: ◐ 18: ● 26: ◑



Tidsmessig isomorfi

Utviklingen av versjonene må gjenspeile virkeligheten på en hensiktsmessig måte

- Hvilken type tidspunkter tidsstempelet skal inneholde, må velges etter hva modellen skal reflektere
- Versjonene må registreres i databasen slik at forsinkelsen fra hendelse til registrering er akseptabel

Hva skal tidsstempelet reflektere?

1. Når en hendelse faktisk inntraff?
2. Når versjonen ble lagt inn?
3. Når versjonen skal tre i kraft?
4. Når versjonen ble ugyldig?
5. ...

Maks én av disse! (men neppe nr. 4)

Merk:

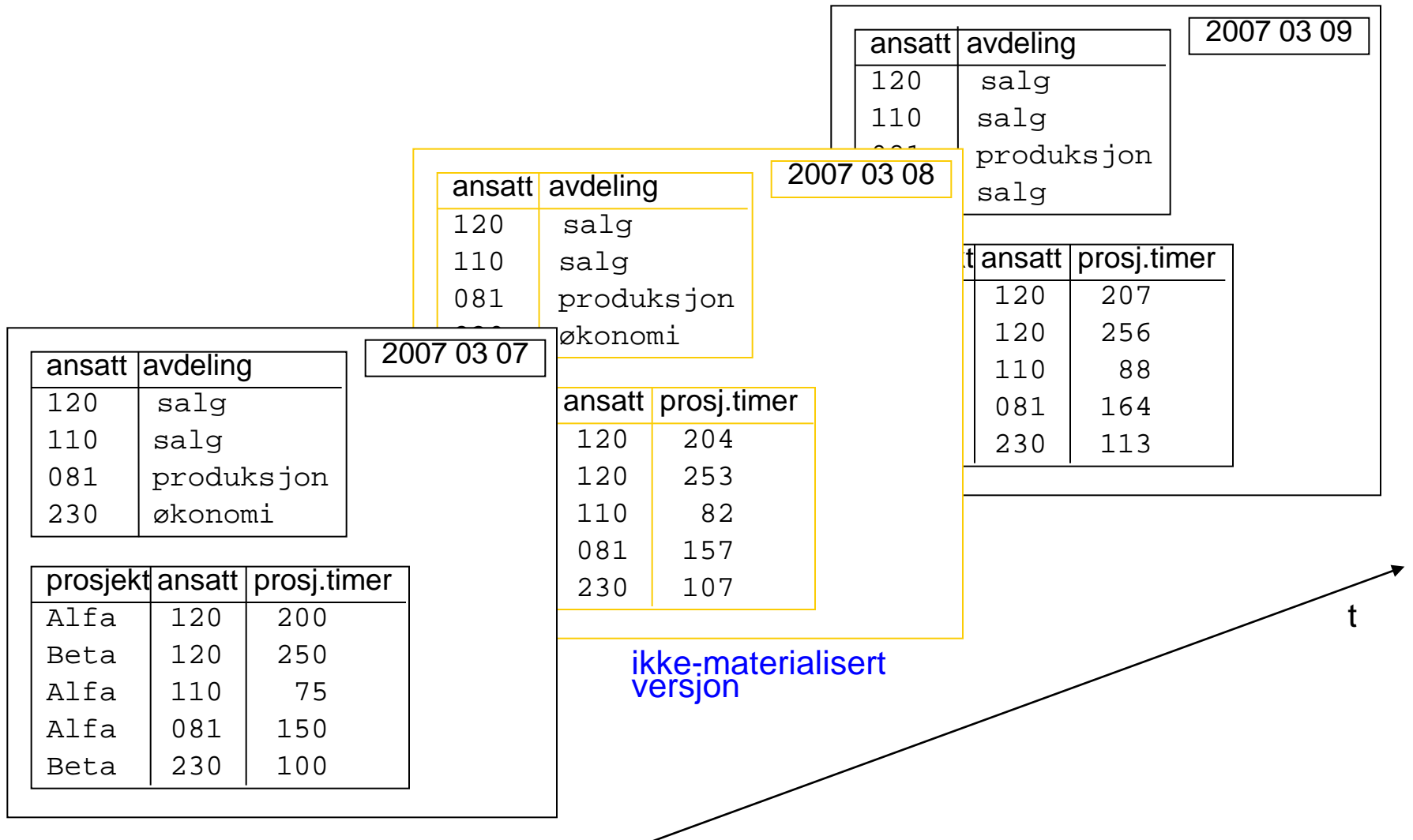
- De fleste modeller ligger *etter* virkeligheten, f.eks. nr.1 (og nr. 2):
Det tar tid før en hendelse i virkeligheten kan gjenfinnes som en versjon i databasen (mikrosekunder til dager, avhengig av registreringsprosess)
- Noen modeller må ligge «foran» virkeligheten, f.eks. nr.3

Tidsmessig kontinuitet

Det er maksimalt én versjon pr. mulig tidsstempel

Dersom det legges inn **færre** enn en versjon pr. mulig tidsstempel, så må det være mulig å avlede **ikke-materialiserte versjoner** for de tidsstemplene som ikke har en tilhørende versjon

Tidsmessig kontinuitet



Film- eller lysbildeprinsippet?

| Vare (.nr) | Dag (Dato) | Beløp (NOK) |
|------------|---------------|-------------|
| vare | gyldighetsdag | listepris |
| såpe | 01.01.07 | 97,- |
| såpe | 02.01.07 | 97,- |
| såpe | 03.01.07 | 97,- |
| ... | | |
| såpe | 01.07.07 | 105,- |
| kost | 01.01.07 | 32,- |
| kost | 02.01.07 | 32,- |
| ... | | |

filmprinsippet

| Vare (.nr) | Dag (Dato) | Beløp (NOK) |
|------------|--------------------|-------------|
| vare | ikrafttredelsesdag | listepris |
| såpe | 01.01.07 | 97,- |
| såpe | 01.07.07 | 105,- |
| kost | 01.01.07 | 32,- |

lysborneprinsippet

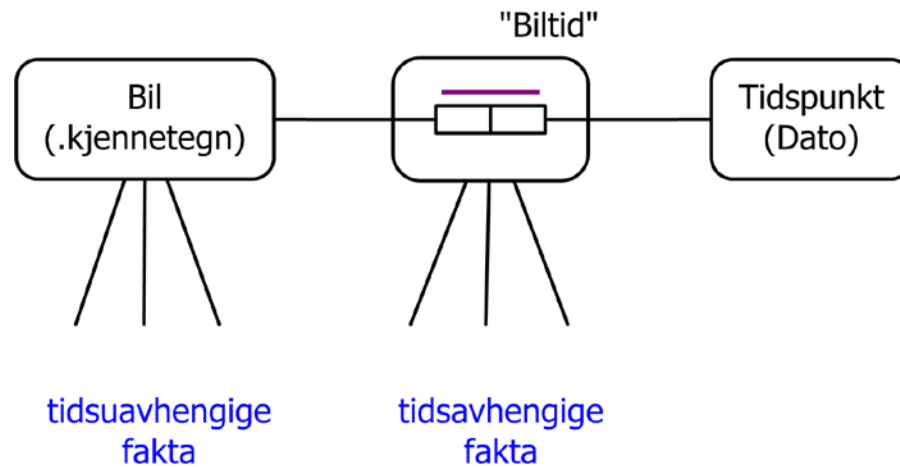
Film- og lysbildeprinsippet

- **Filmprinsippet:**
 - Én ny versjon for hvert nytt mulig tidsstempel
 - Trenger mye lagerplass ved fin granularitet
- **Lysbildeprinsippet:**
 - Observerer og registrerer virkeligheten bare av og til
 - Bygg inn nok kunnskap til at de ikke-materialiserte versjonene kan utledes

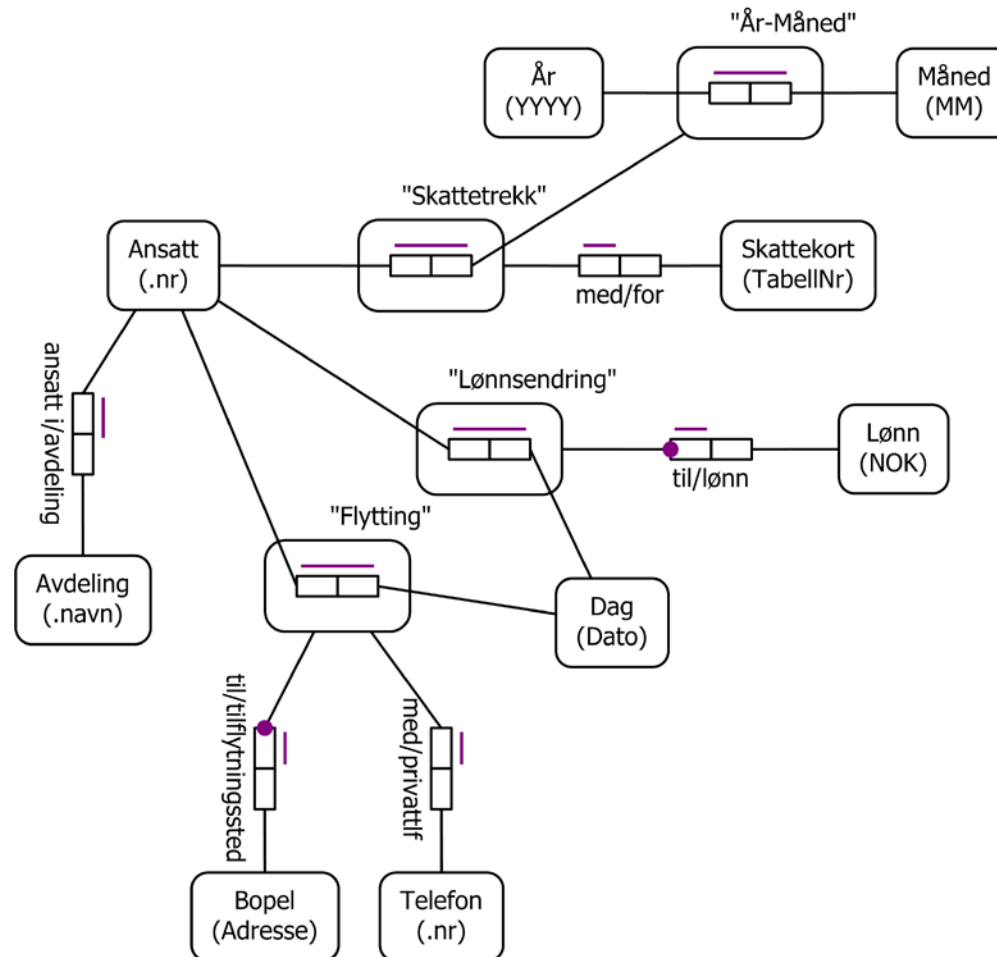
Elementære setninger og tid

- Tre typer elementære setninger:
 - De som ikke har noen tidsdimensjon
 - De hvor vi bare ønsker å ta vare på siste aktuelle verdi
 - De hvor vi ønsker å modellere en tidsdimensjon
- Virkeligheten har to typer endringer:
 - Kontinuerlige
 - Sprangvise
- Virkelighet kontra modell:
 - Versjonene endrer seg alltid i rykk og napp
 - Versjonene kan være tidsmessig forskjøvet i forhold til virkeligheten

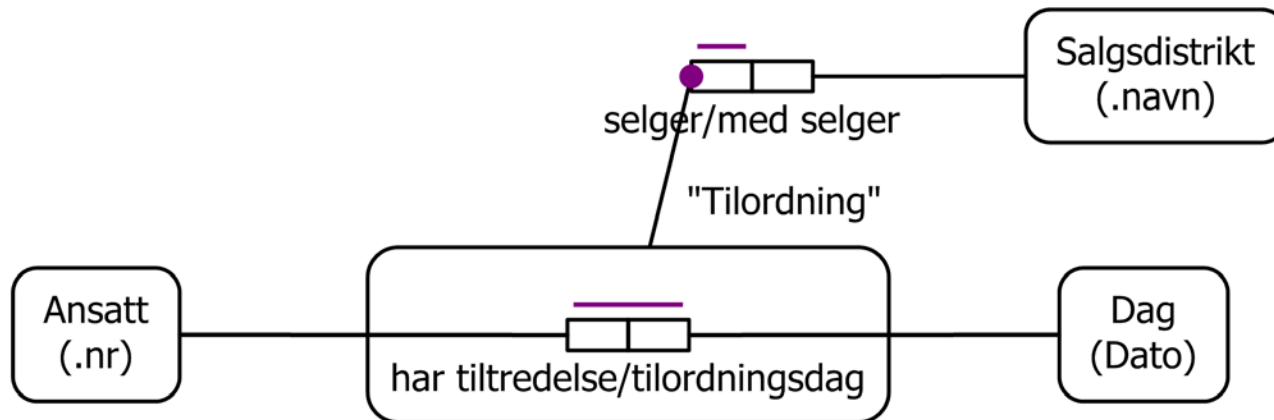
Begrepsdannelse med tidsaksen



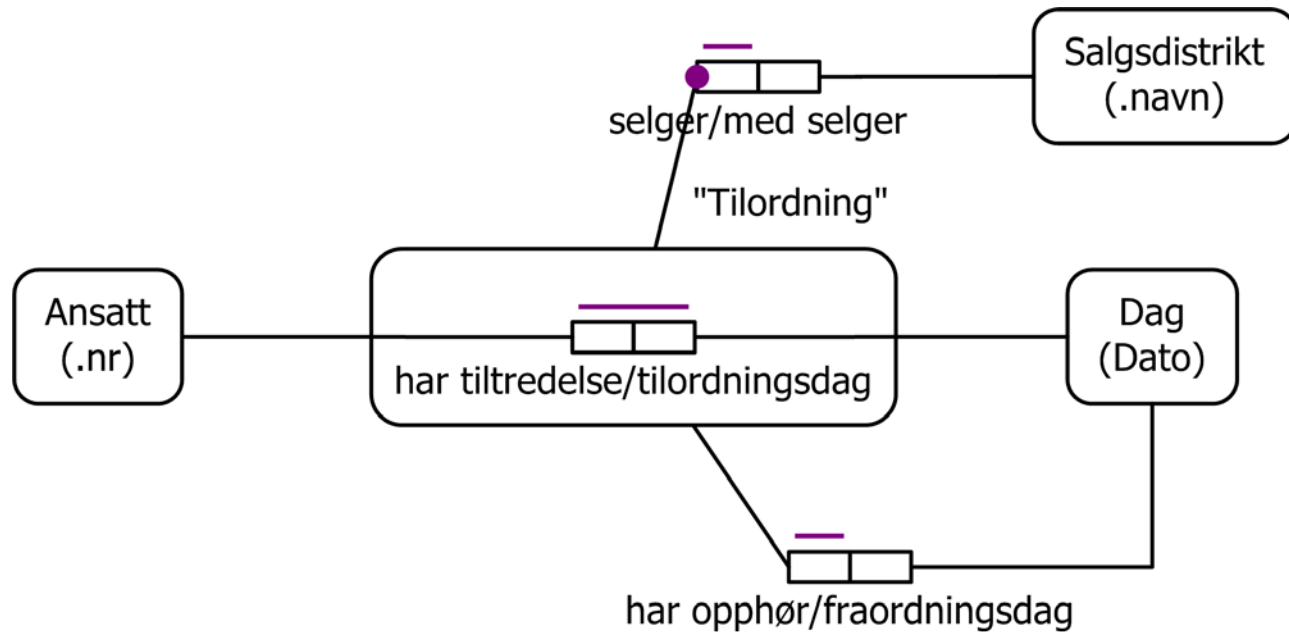
Eksempel



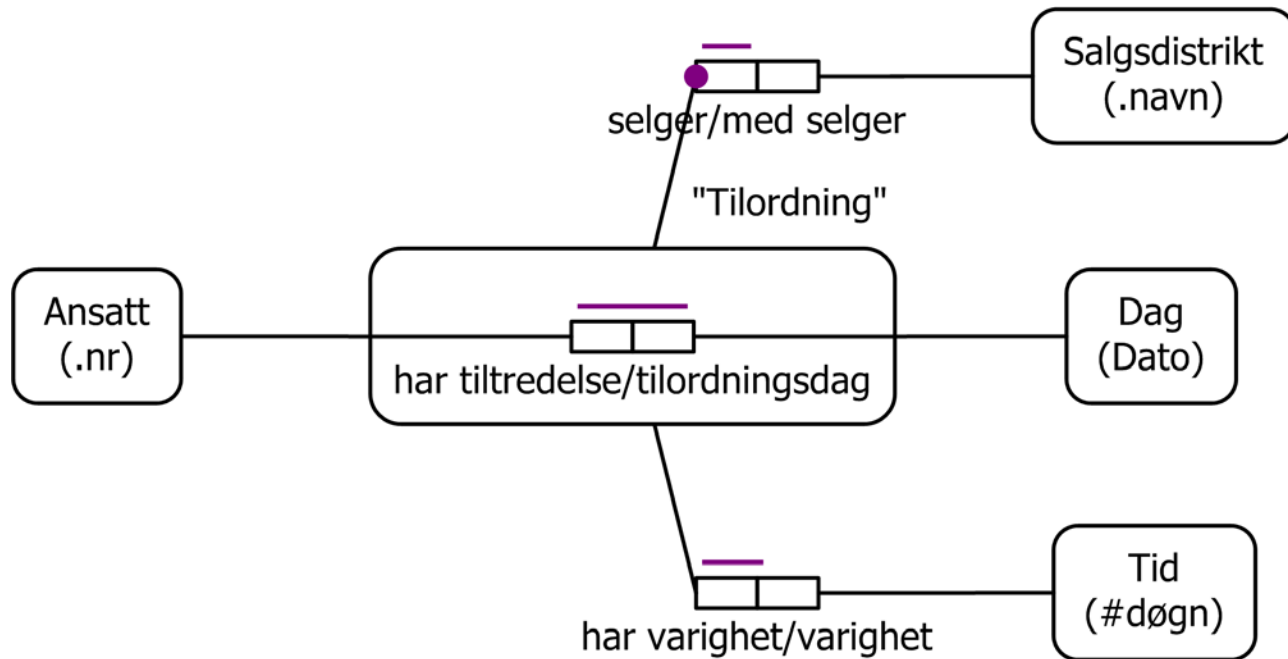
Eksempel: Hvor lunge varer det?



Angivelse av opphør



Angivelse av varighet



Representasjoner

- Alle begreper må kunne representeres
 - Begrepsforekomster kan ikke lagres; det vi lagrer, er *representasjonsforekomster*
- Skal vi kunne realisere modellen som en relasjonsdatabase, må vi representere alle begrepene entydig

Valg av representasjon

- Valg av representasjon:
 - entydig bro mellom en forekomst av en representasjon og forekomsten av det tilhørende begrepet
 - helst uforanderlig
 - støtte utveksling av informasjon mellom systemer
- **Identifikator** = representasjon hvor det er en uforanderlig en-til-en-bro mellom begrep og representasjon

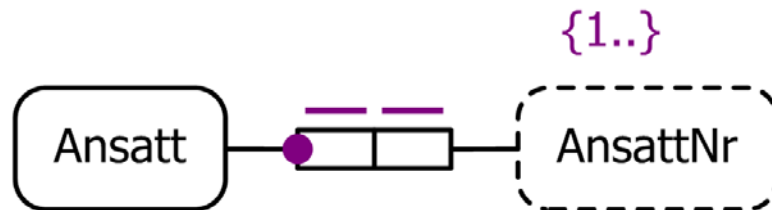
Representasjonstyper

- navn, koder, forkortelser
- boolske verdier
- tellbare størrelser
- tids- og romlige verdier
- fritekst
- representasjoner av bilde og lyd

Ikke-informasjonsbærende representasjoner

Representasjonen til begrepet identifiserer en forekomst av begrepet

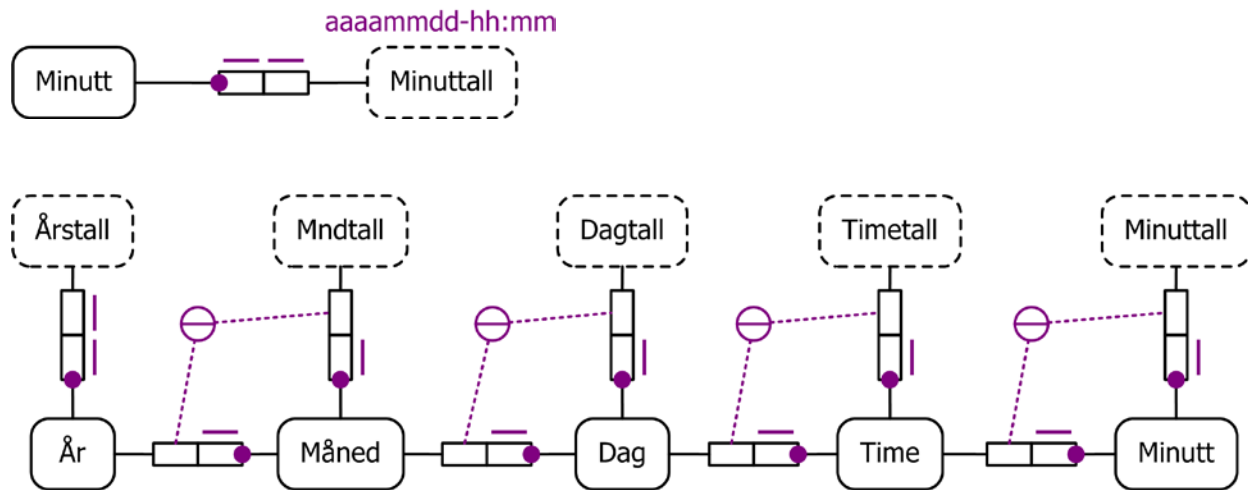
Det fins ingen innkodet informasjon i representasjonen



Delvis informasjonsbærende representasjoner

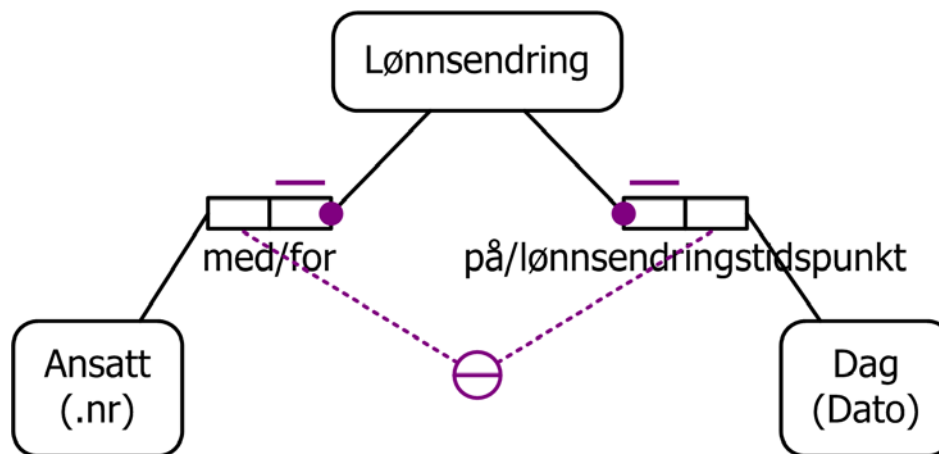
Deler av representasjonen til et begrep identifiserer en forekomst av et annet begrep

Det kan, men behøver ikke, være synlig i modellen



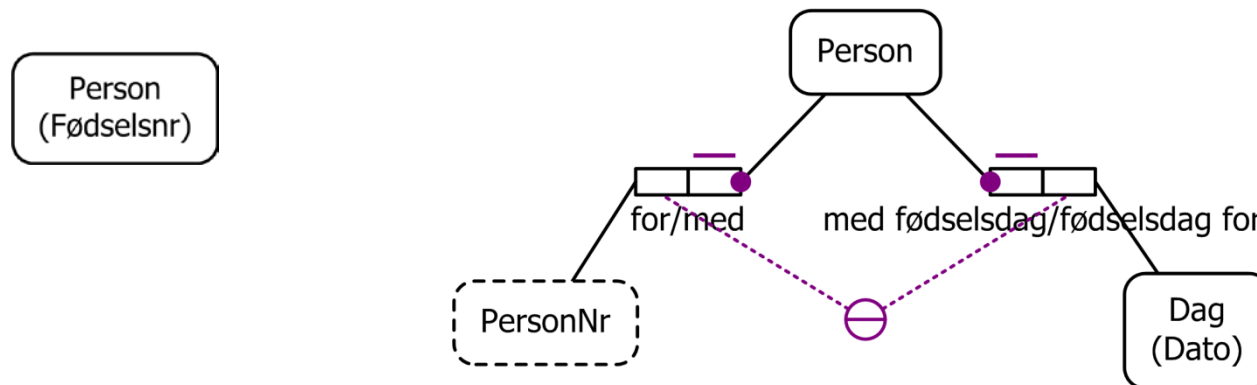
Totalt informasjonsbærende representasjoner

Representasjonen til begrepet består av en samling elementer der hvert element identifiserer en forekomst av et annet begrep



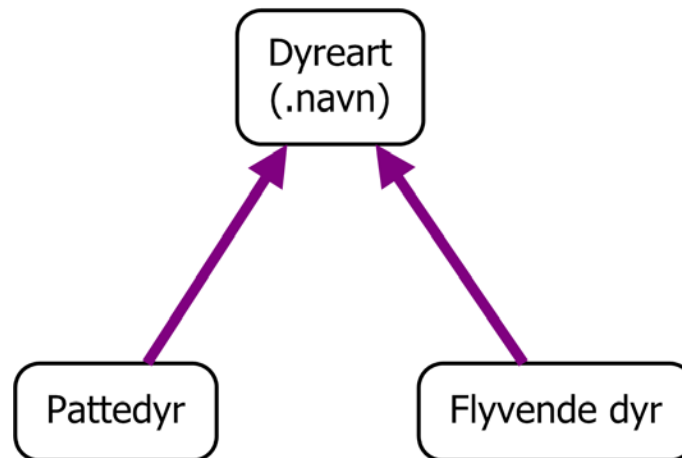
Synliggjøring eller ikke av informasjonsbærende representasjon i modellen?

Hvis det er en mulighet for at brukeren etterspør denne informasjonen, bør den vises i modellen



Representasjon via superbegrep

Underbegreper arver alltid representasjonen til sitt superbegrep



Realisering av underbegreper

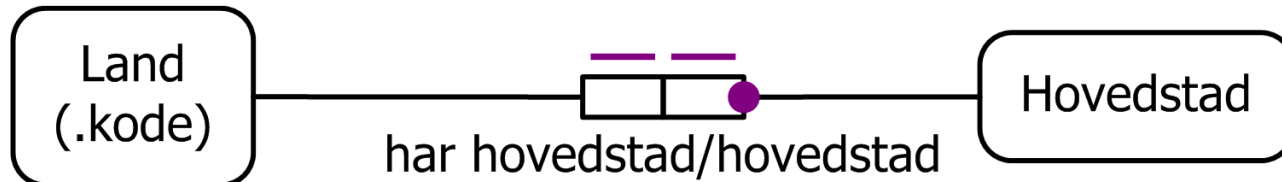
- Antar:
 - Alle superbegrep har en underbegrepsforklaring
 - Alle underbegrep har minst en gruppererolle

Til hvert underbegrep opprettes en relasjon med samme primærnøkkel som superbegrepet, og med fremmednøkkel fra underbegrepets primærnøkkel til superbegrepets.
- To alternativer:
 1. Attributter fordeles mellom superbegrepets og underbegrepets relasjoner
 2. Alle attributtene fra superbegrepets relasjon (unntatt underbegrepsforklaringen) gjentas i underbegrepets relasjon

Representasjon via en-til-en-faktatype

Et begrep med en påkrevd rolle i en en-til-en-faktatype til et annet begrep kan identifiseres indirekte gjennom det andre begrepet

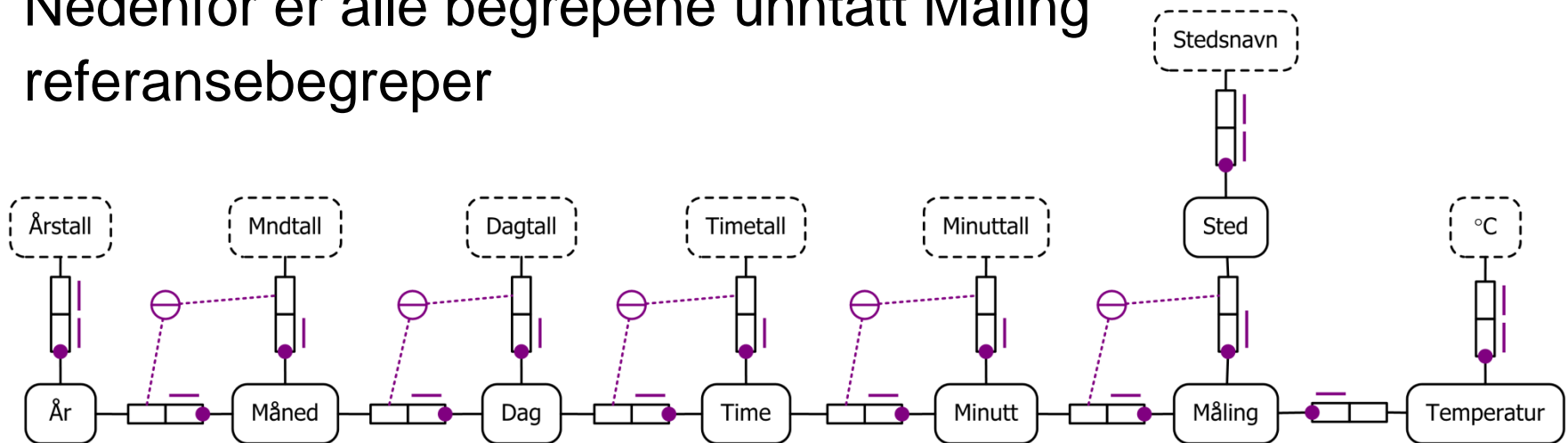
Eksempel: En hovedstad kan identifiseres med det landet den er hovedstad i



Undertrykking av begreper I

Et begrep som i realiseringsalgoritmen ikke spiller noen andre grupperoller enn de som inngår i den prefererte referansen, og som spiller minst en referanserolle, kalles et **referansebegrep**.

Nedenfor er alle begrepene unntatt Måling referansebegreper



Undertrykking av begreper II

Tabeller som kommer fra referansebegreper, kan **fjernes** (**undertrykkes**) fra relasjonsdatabase-skjemaet.

Noen ganger ønsker vi likevel å beholde slike tabeller, f.eks. for å bevare informasjon som ellers ville gått tapt når forekomster i andre tabeller slettes.

Et eksempel fra figuren på forrige lysark, er tabellen Sted: Hvis vi undertrykker Sted, og alle målinger for Blindern fjernes, mister vi informasjon om at det fins et sted som heter Blindern.