




UNIVERSITETET  
I OSLO

# Datasystemer og informasjonssystemer

Forelesning, DRI1001, 14. september 2010



UNIVERSITETET  
I OSLO

## Disposisjon

- Hva er en datamaskin og hva kan den brukes til
- Prinsipper for datamaskinens virkemåte
- Data, informasjon og kunnskap
- Formalisering
- Datasystemer og deres omgivelser
- Informasjonssystemer

## Datamaskin

- Definisjon (Norsk dataordbok):

Funksjonsenhet som uten menneskelig inngripen under selve operasjonen kan utføre omfattende beregninger med et stort antall logiske og aritmetiske operasjoner

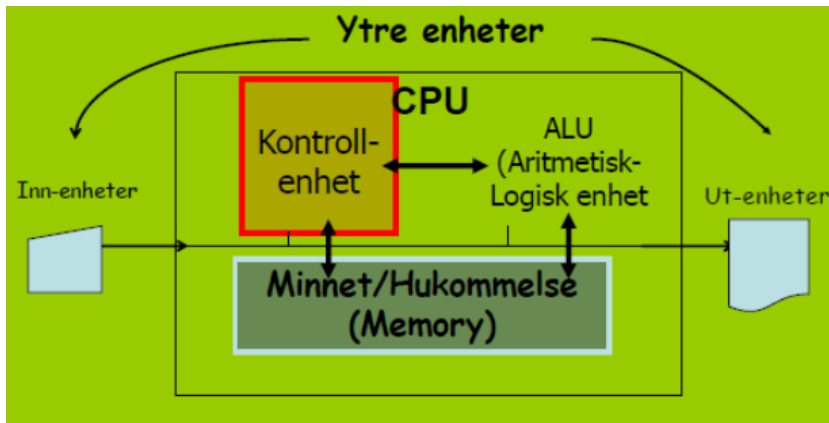
En datamaskin kan ha en eller flere prosessorenheter og ytre enheter. Den kan være frittstående eller bestå av flere sammenkoblede enheter

- Grei definisjon, men mangler kanskje noe om "data"?

## Maskinarkitektur, von Neumann

- Minnet
    - "Primærlager", der data og programmer lagres under programutføringen
  - Aritmetisk-logisk enhet (ALU)
    - Her utføres programmet
  - Kontrollenheten
    - Kontrollerer og styrer de øvrige enheter
- } CPU
- Inn-enheter
    - Forer maskinen med data som skal behandles (tastatur, disk ...)
  - Ut-enheter
    - Presenterer resultatet av behandlingen (skjerm, disk, papir ...)

## Von Neumann, prinsippskisse



## Noen av datamaskinens karakteristiske trekk

- Representasjoner av data
  - Digitalisering
  - Formalisering
- Programmerte operasjoner
  - Nøyaktighet, presisjon
  - Forutsigbarhet og repeterbarhet
- Ytelse
  - Kontinuitet, døgnet rundt, ingen "blåmandager"
  - Stadig mer komprimert, høy hastighet og lagringskapasitet
  - Stabil og robust

## Digitalisering (ment som en liten avmystifisering...)

- "Digital" kan oversettes til "sifferbasert" (digit = siffer)
- Datamaskinens digitale representasjoner er grunnleggende binære
  - To-tallssystem, dvs. hvert siffer kan ha to verdier (bi- = to-)
  - En BIT (står for "Binary digiT") har verdien 1 eller 0
- Lite eksempel:
  - Tallet 123 i titalssystemet:  $1 * 10^2 + 2 * 10^1 + 3 * 10^0 = 100 + 20 + 3$
  - Tallet 101 i to-tallssystemet:  $1 * 2^2 + 0 * 2^1 + 1 * 2^0$   
(de tre sifrene 101 tilsvarer 5 i ti-tallssystemet, "4 + 0 + 1")
- Hvorfor to-verdi representasjon?
  - Teknisk: Strøm/ikke-strøm
  - Logikk: Egnert til å representere dikotomien "sann/usann"

## Digitalisering av tegn, eller "tekst"

- Representasjon av de enkelte "tegn": Bytes
  - 1 byte = 8 bits, gir  $2^8 = 256$  kombinasjonsmuligheter
  - Særlig innflytelsesrik standard: ASCII (American Standard Code for Information Interchange), bruker de første 128 sifrene
  - Eks.:
    - ! = tegn nr. 33
    - 0 = tegn nr. 48
    - 3 = tegn nr. 51
    - A = tegn nr. 65
    - a = tegn nr. 97 (etc.)
  - De neste 128 sifrene ble ofte brukt til forskjellige "lokale utvidelser" (vesteuropeisk hadde et annet oppsett enn sentraleuropeisk etc.)
  - De seneste årene: Standardisert, UTF (Unicode Transformation Format) gir et felles tegnsett av "nesten ubegrenset størrelse"

## Dataformater for annet enn "tekst"

- Bilder
  - **Punktgrafikk** (hvert punkt i et areal har en tallverdi som representerer fargen)
  - **Vektorgrafikk** (bildet er representert som opplysninger om hva slags geometrisk "primitiv" som skal plasseres mellom bestemte koordinater i bildet)
- Musikk
  - **Audio** (representasjon av lydbølgene, eks. wav-filer, MP3 osv.)
  - **MIDI** (informasjon om hvilket instrument, rekkefølge og varighet av hver tone, anslagstype etc. Dermed kan maskinen "spille noter")
- Representasjoner på disse basale nivåene er lite relevant for DRI! Men *nå* kommer det, altså... ☺

## Komplekse (tegnbaserte) dataformater

- Sekvenser av tegn
  - Fri tekst, minimal struktur, "kan representeres på papir"
  - "Hei", "Barack Obama", "14.07.1789", "J45"
- Formater og "metadata"
  - Informasjon om "meningen" flyttes fra leserens hode til maskinen
  - Formatene  $\langle \text{fornavn} \text{ mellomrom } \text{etternavn} \rangle$ , *dato* på formen  $\langle \text{dd.mm.åååå} \rangle$ , *Koden* for "astma" i diagnosekatalogen ICD-10
- Datastrukturer
  - Begreper sammenheng: person  $P$  er president i landet  $L$
  - Kompleks "grafstruktur": Flyplasskoder, og flytiden mellom dem  
OSL  $\rightarrow$  CPH, 1:10. OSL  $\rightarrow$  CDG, 2:20. CPH  $\rightarrow$  CDG, 2:00.

## Formalisering av data

- Ikke naturgitt hvordan "ting i verden" skal representeres som tekst, koder, datastrukturer etc.
- Hvem bestemmer hvilke sykdomskoder som skal brukes?
  - WHO (enhet i FN-systemet) vedtar kodeverket ICD
  - Endringer er underlagt et omfattende "forhandlingsmaskineri"
    - Ulike oppfatninger av hvilke "klasser" en sykdom skal innordnes under
    - Ulike oppfatninger av hvilke kjennetegn som er mest vesentlige for at en bestemt type diagnose skal stilles
  - Norge (og andre land) bestemmer selv om de skal følge kodeverket

## Formalisering – hvordan identifisere noe?

- Offentlig saksbehandling: Saksnummer
  - Ofte et nummer basert på årstall og femsifret løpenummer
  - Eks.: 2007/01982
  - Hvilket organ hører denne saken hjemme i?
- Lokale identifikatorer, krever sammenhengskunnskap
  - Hvem bruker identifikatoren?
  - Brukes identifikatoren til alle typer saker, eller bare bestemte typer?
  - Hva er kriteriet for å bestemme om et brev hører til i den ene eller i den andre saken?
- Universelle identifikatorer, må forvaltes av noen
  - Personnummer: Skattedirektoratet
  - Foretaksnummer: Brønnøysundregistrene

## Formalisering – hvordan identifisere noe?

- Personnummeret – et barn av sin tid (kom i 1964)
- 29027541793
  - 6 første sifre: dato (to-sifret årstall)
  - 7. siffer: over eller under 5: indikerer århundret
  - 7-9. siffer: Løpenummer
  - 9. siffer: partall for kvinne, oddetall for mann
  - 10. siffer: Kontrollsiffer som beregnes ut fra de 9 foregående
  - 11. siffer: Kontrollsiffer som beregnes ut fra de 10 foregående
- Altså: Max 250 personer av hvert kjønn hver dato
- Utlendinger i Norge: D-nummer (første siffer: +3)
- Man ville neppe designet nummeret slik i dag...

## Formalisering – kan både omfatte *dataene* og *behandlingsprosessene*

- Eks. søknad om en ytelse fra forvaltningen:
  - Skal ofte sendes "på fastsatt skjema" (papir eller elektronisk)
  - Krever at bestemte rubrikker skal fylles inn
  - I elektronisk saksbehandling settes ofte dine utfylte opplysninger sammen med andre opplysninger som er lagret på forhånd
    - Lånekassen vet fra før hvor mange studiepoeng du har
    - Skatteetaten har fått inn de fleste opplysningene om dine inntekter, lån, formuesgjenstander osv. fra ulike hold
  - Forvaltningens vedtak skal bygge på *dataene* og tolkning av *regler*
  - *Informasjonsbehovet* bygger på en formalisering
  - Kunnskap om hvilke data som er tilstrekkelige til å oppfylle informasjonsbehovet er en formalisering
  - Regne- og slutningsreglene (for å treffe vedtak) er en formalisering

## Datamaskinprogrammer

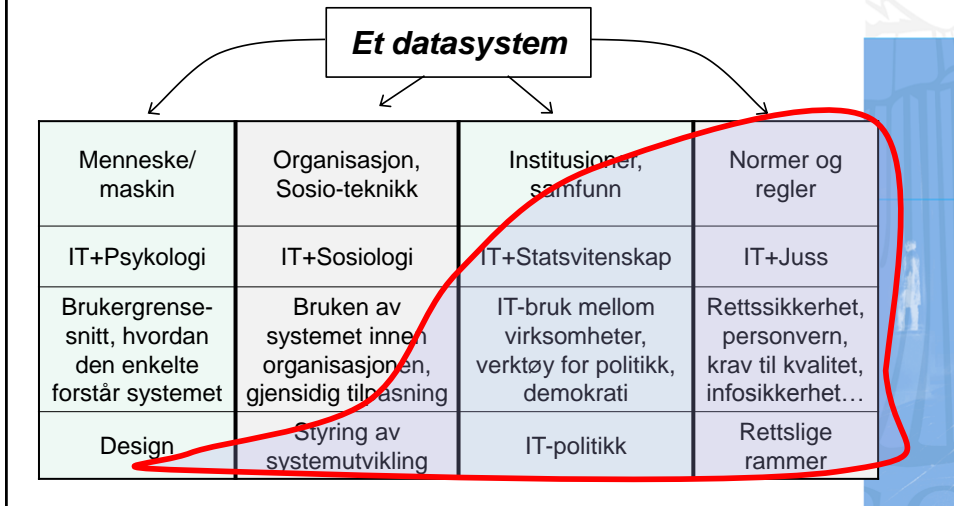
- Fra definisjonen i begynnelsen av forelesningen:  
*utføre omfattende beregninger med et stort antall logiske og aritmetiske operasjoner*
- Programmering dreier seg altså om *operasjoner*.  
Hva maskinen gjør, og hvordan den gjør det
- På samme måte som data kan betraktes i ulike nivåer, er også programmene delt inn i ulike nivåer
  - Maskinkode (binærnivået)
  - Operativsystem (styrer ressurser og tilkoblede enheter etc.)
  - Applikasjonsprogrammer (programmene vi "bruker")
  - Systemintegrasjon (samspill mellom applikasjoner)

## Begrepet "datasystem"

- Forholdsvis vidt begrep, men det kan avgrenses til håndfaste, konstruerte, og utskiftbare komponenter
  - I praksis bruker vi det om *elektroniske datasystemer*, men i prinsippet kan også manuell bokføring være et "datasystem"
- En definisjon (blant flere mulige):  
*Et system for innsamling, bearbeiding, lagring, overføring og presentasjon av alle former for data*
- *Datasystemet* er resultatet av formaliseringer, tolkninger, valg, forhandlinger, beslutninger, innkjøp, programmering etc.
  - Likevel er perspektivet at *et veldefinert problem har én beste løsning*



## Datasystemet og dets omgivelser



## Menneske/maskingrensesnitt

...et tema vi i beskjeden grad berører i DRI-kursene  
(mest som "universell utforming", i en IT-politikk/rettslig ramme-kontekst)



Ikke nødvendigvis bare skjerm, tastatur og mus...

## Informasjonssystem

- En ganske brukbar definisjon:
  - Samlingen av menneskelige og maskinelle ressurser samt regler og prosedyrer organisert for å utføre bestemte funksjoner og løse en bestemt oppgave*
  - Definisjonen er ikke veldig presis
  - Bør kanskje ikke være det heller?
  - Ordet informasjonssystem brukes på mange forskjellige måter
- I en eller annen forstand dreier det seg om å forstå et datasystem i sammenheng med omgivelsene
  - Men ikke "alt som finnes i omgivelsene" inngår i *informasjonssystem*
  - For eksempel er det lite naturlig å trekke inn regelverk for offentlige anskaffelser i forståelsen av hva et informasjonssystem er

## Noen syn på det å utvikle IS

- Konstruksjonsprosess
  - Utvikle en ny teknisk løsning, som et verktøy (f.eks. implementere en endring i regelverket; eks. NAVs pensjonsprosjekt)
- Teknisk og organisatorisk forandringsprosess
  - Studentweb, elektronisk ligning, Lånekassas web-baserte søknadsbehandlersystem
- Erkjennelsesprosess
  - Analyse av organisasjonen: f eks. ligningsetatens FLID-prosjekt
- Politisk prosess (kommer "utenfra" organisasjonen)
  - Felles arkitektur/felleskomponenter
  - AltInn, MinSide

## Andre perspektiver på IS

- Et fortolkende perspektiv, viktig for å kunne forstå virkninger (og ikke-virkninger) av teknologiendring
  - Hvorfor lykkes noen prosjekter?
  - Hvorfor går noen prosjekter galt?
  - Svarene finner man til dels i alle de fire kategoriene av omgivelser
- Rammene rundt systemene er ikke stabil, mål og forventninger endrer seg ofte underveis
  - Sak styring innad i organisasjonen?
  - Eller forhold utenfor som man bare har begrenset kontroll med?
- Komplekst samspill mellom stabilitet og endring
  - Gammel "systemarv" kan være både til fordel og til ulempe når man vil endre noe

## Sammenligning DS / IS

### Datasystem

- Kan være flere innen samme IS
- Et veldefinert problem har "én beste løsning"
- DS omgivelser er et avgrenset felt, som man skal kunne håndtere optimalt i en systemutviklingsprosess

### Informasjonssystem

- Omfatter mer enn bare det som finnes i ett (eller flere) DS
- Aldri entydig hva som er "beste løsning"
- DS omgivelser er i prinsippet grenseløse: IS vedrører utvalg, tolkning og håndtering av ulike sider ved omgivelsene