

HØGSKOLEN I OSLO  
OG AKERSHUS

---

## *Kompetansebehov for fremtidens arkivarer*

– Hva skal til for å bruke digitalt skapt arkivmateriale i depotene?

Arkivarforeningens vårseminar, 11. april 2018

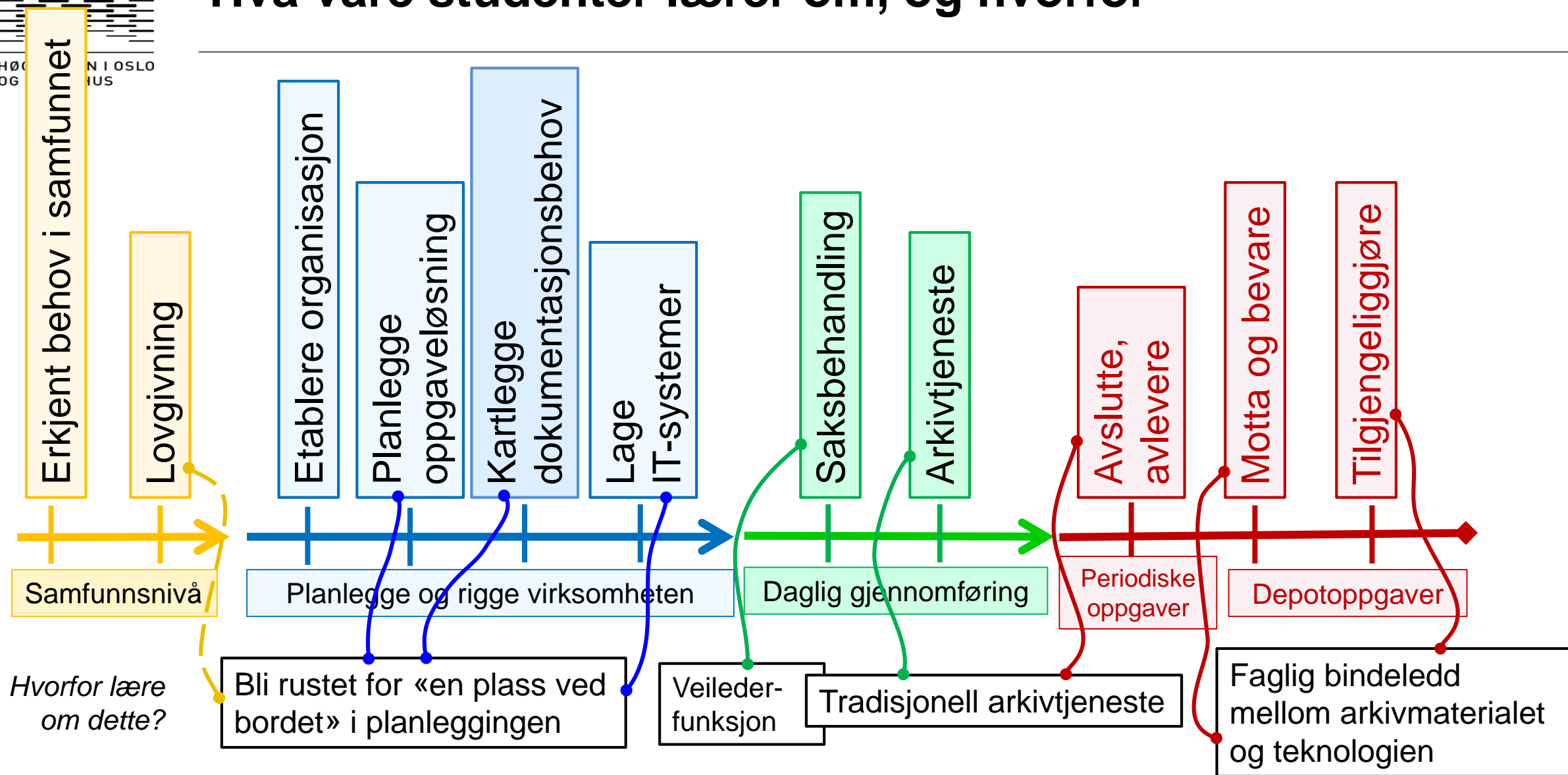
Herbjørn Andresen  
Førsteamanuensis

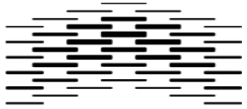
OsloMet – storbyuniversitetet  
Institutt for arkiv-, bibliotek- og informasjonsfag



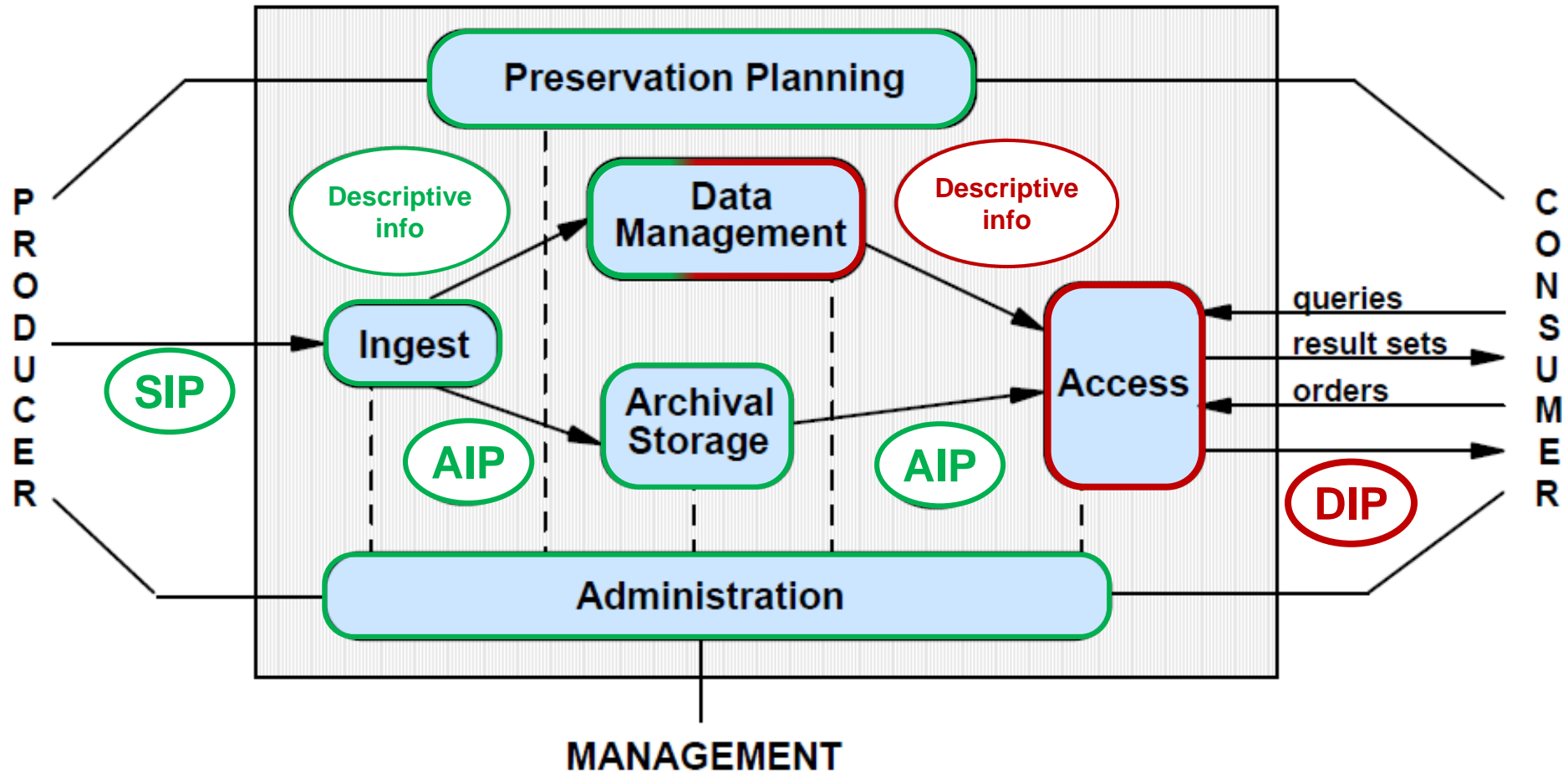
# Hva våre studenter lærer om, og hvorfor

HØGSKOLEN I OSLO  
OG BIENNALETT  
US





# «Gammelt» og «nytt» digitalt depot-emne

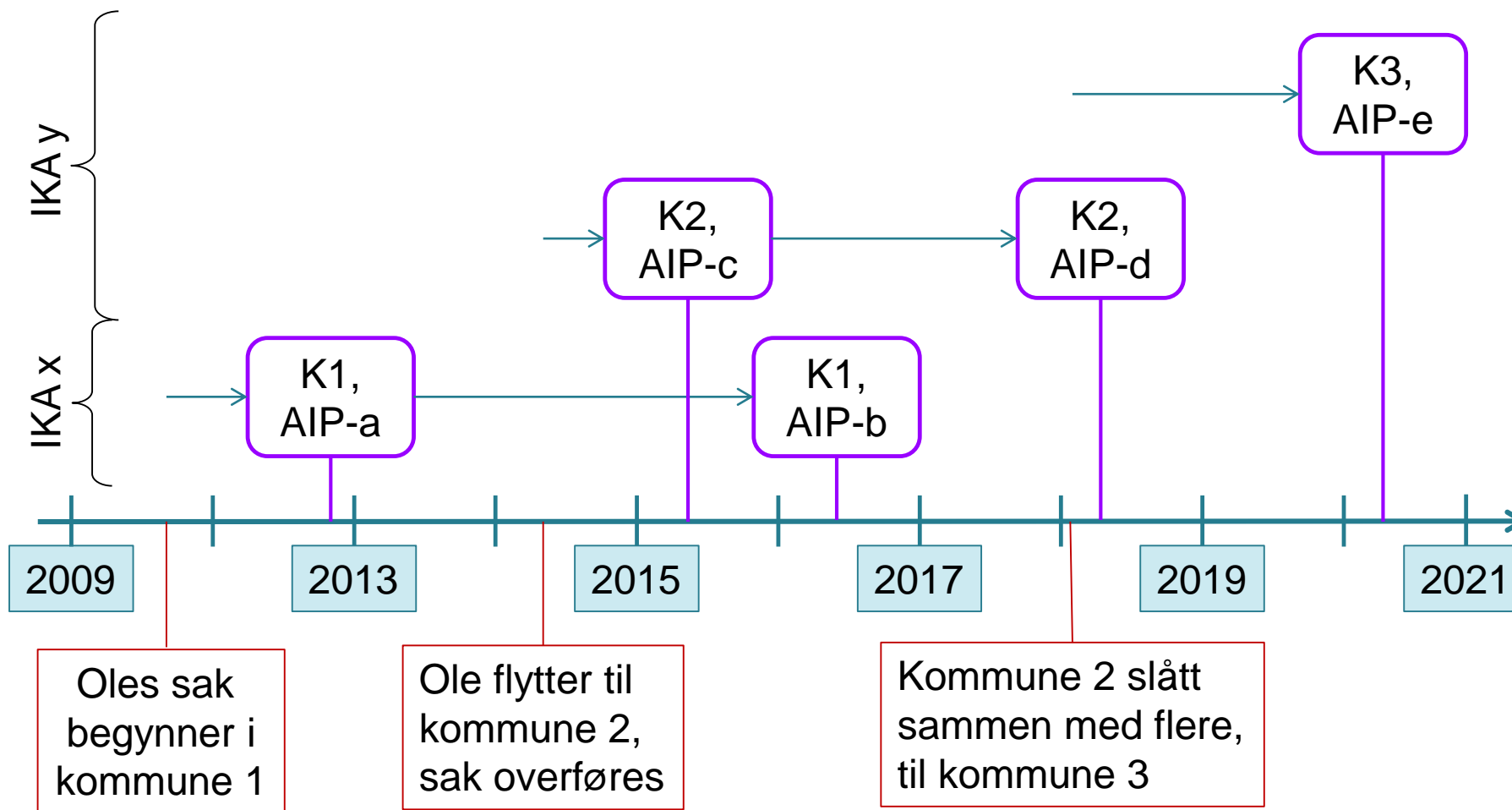


## Hva befinner seg i et digitalt depot?

- Arkivpakker («AIP», Archival information packages)
  - Bevaringsobjektene, det materialet som skal holdes forståelig og autentisk over tid, og gjennom ulike teknologiske endringer
- Arkivbeskrivelser (Descriptive information)
  - Informasjon som kan brukes til å finne frem i arkivpakkene, kan bearbeides og suppleres etter behov
- Formidlingspakker («DIP», Dissemination information packages)
  - Innholdet kan være identisk med én AIP, eller være bearbeidet og satt sammen av (deler av) flere AIP-er
  - Ikke nødvendigvis bevaringsobjekter, beholdes så lenge det er hensiktsmessig

## Eksempel – en DIP for Oles PP-tjenestekarriere 2009-2019

*Periodiseringer, flytting og organisasjonsendringer sprer informasjonen til ulike AIP-er (som i et papirarkiv, men lange forløp flises mer opp på grunn av hyppigere periodisering)*



## Forskjellige kvalitative egenskaper ved digitalt arkivmateriale

- En del registre er designet for negativ troverdighet
  - Kvalitet = samsvar mellom data og verden
  - Folkeregisteret, Enhetsregisteret, Autosys
- Dokumentasjon organisert i en arkivstruktur
  - Kvalitet som autentisk uttrykk for handlingene som er dokumentert
  - Sammenhengene har informasjonsverdi, mulig å tolke «meaningful absences»
- Dokumenter og data som søkes opp uavhengig av arkivstrukturen
  - Kvalitet vurderes ut fra egenskaper ved det enkelte dokument, diplomatikk
  - Eventuelle sammenhenger kan undersøkes «bottom up»

## Hvordan finne frem – hvordan koble sammen

- Koble sammen = lage DIP-er
  - Noenlunde automatiserbart der AIP-er er basert på samme Noark-versjon, eller basert på fagsystemer som har samme datasettbeskrivelse for tabelluttrekk
  - I andre situasjoner vil det være mye bearbeiding og vurderinger, kan forhåpentligvis bli overkommelig med gode støtteverktøy
- Finne frem, ulike strategier
  - a) Bygge og bruke arkivbeskrivelser
  - b) Information retrieval, «fritekstsøk» i dokumenter og metadata
- Noen fordeler og ulemper ved de ulike strategiene



# Status pr. i dag – relativt tynt med arkivbeskrivelser for digitalt skapt materiale

## Et eksempel fra Arkivportalen, forholdsvis gammelt

Arkiv: A-100729 - Sentralskattekontoret for utenlandssaker

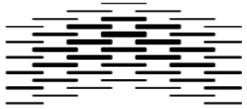
Aktør: A-100729 - Sentralskattekontoret for utenlandssaker

Nummer: Innhold:	Tidsrom
<b>A-100729 Sentralskattekontoret for utenlandssaker</b> Uttrekk av database over skatteyttere, personer og bedrifter 1984-2001. Systemnavn: SFUFILER - Egenutviklede fagsystemer. Uttrekksmetode: Komplette registre pr. 04.01.2001. Startdato for uttrekket:01.01.1984. Sluttdato for uttrekket:04.01.2001. Merknader: Inneværende år"=2001 mht. skattemantall etc. Avleveringen omfatter følgende CD-R: ARKIV1. Metadata, dokumentasjon, systemets tabeller	<b>1984 - 2001</b>







# Mer vanlig foreløpig, elektronisk journal som følger med papirarkiv (Eksempel fra en Noark 4-journal)

- [-] SKOLE-036-01 Høgskoleadministrasjonen, Høgskolen i Hedmark, 1994 -
- + [ ] A Møtebøker, referatprotokoller, forhandlingsprotokoller o.l.,
- + [ ] B Kopibøker,
- [ ] **C Journaler og overgripende registre,**
  - + [ ] Ca Postlister, 1994 - 1995
  - + [ ] Cb Journaler, 1995 - 1999
  - [ ] Cc Arkivuttrekk fra elektronisk journal, 01.03.1995 - 31.12.2004
    - [ ] 0001 Metadata,
    - [ ] 0002 Data,
    - [ ] 0003 Dokumentasjon,
    - [ ] 0004 Rapporter,
- + [ ] D Sakarkiv ordnet etter organets hovedsystem,
- + [ ] G Opptak,
- + [ ] L Årsmeldinger,
- + [ ] R Regnskap,
- + [ ] X Egenproduserte trykksaker,
  - [ ] Z Referansemateriale (arkivlister, instruksjer, avisutklipp, etc.),

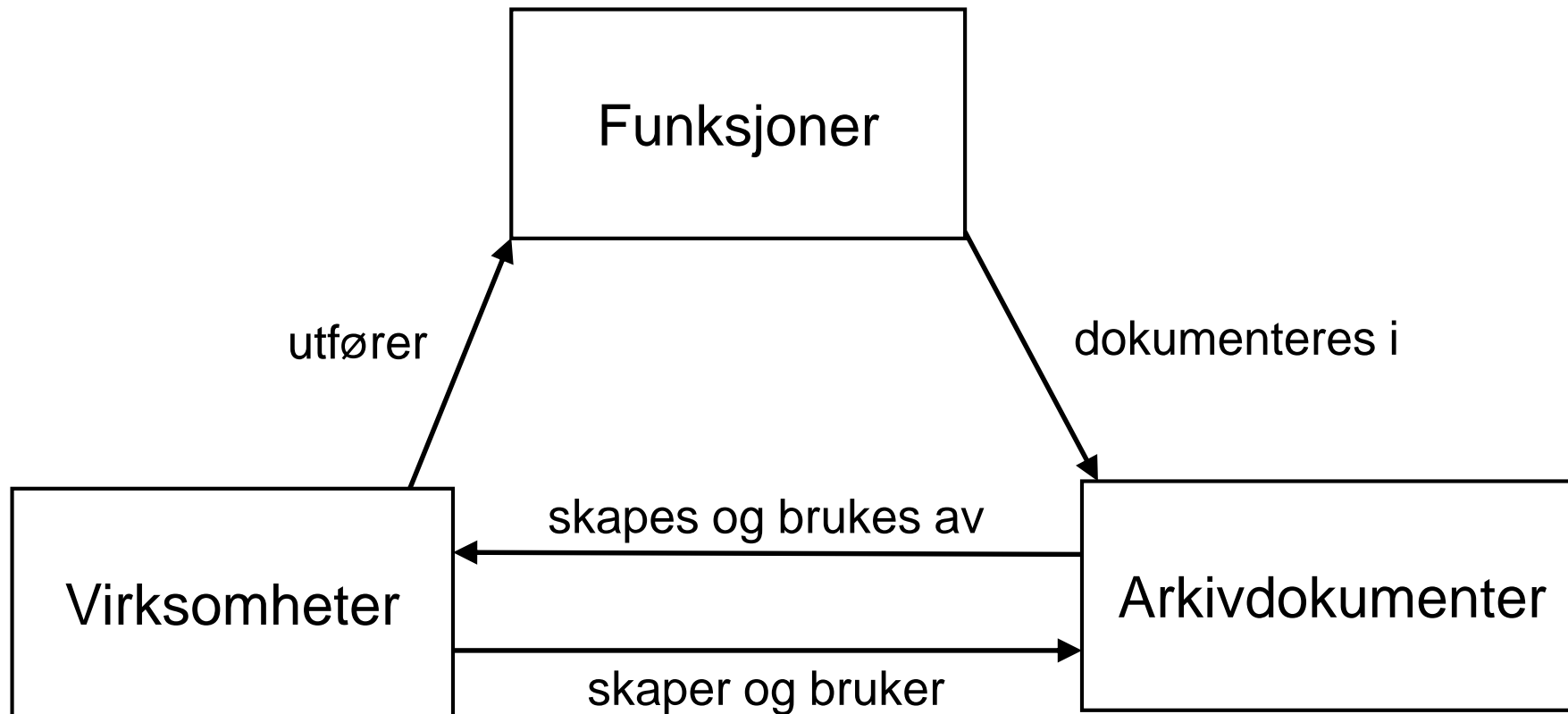


## Hva sier denne beskrivelsen oss?

-  0001 Metadata
  - Omtaler to tekstfiler som beskriver hva som er avlevert
-  0002 Dokumentasjon
  - Testrapport
-  0003 Data
  - Xml-filene med «selve» journalinnholdet
-  0004 Rapporter
  - To obligatoriske rapporter fra Noark-4
- **ALTSÅ:** Det sier omtrent ingenting. Det ligger ikke noe mer informasjon her enn at «noe elektronisk er avlevert»

# Arkivbeskrivelse kan gjøres bedre

Én mulig tilnærming er R-i-C (Records in Context) ICAs «ontologi» for arkivbeskrivelser



## Å bygge arkivbeskrivelser kan delvis automatiseres

- Noe i toppen av hierarkiene må uansett vurderes og beskrives
- De underliggende detaljene kan kanskje generes automatisk?
  - Det er ikke nødvendigvis opplagt hvilke søkebegreper det er hensiktsmessig å legge ut i katalogen, neppe «one size fits all»
  - Hvis man lager arkivbeskrivelser basert på Noark-strukturen, f.eks. ned til mappetittel, vil man antakelig få ganske god «aboutness»-informasjon, men dårlig med «ofness»-informasjon
- Det er i prinsippet ingenting i veien for å lage flere alternative arkivbeskrivelser til samme arkivpakke

## Hva slags alternativ er *Information Retrieval (IR)*?

- Brukeropplevelse «omtrent som Google»
  - Man søker etter ord og uttrykk, og får returnert en liste over hvilke dokumenter eller datatabeller som inneholder det man søkte etter
- Treffene rangeres, slik at det du mest sannsynlig har vært ute etter kommer høyest opp i trefflisten
  - Om «spesialundervisning» er et tilfeldig ord i et enormt dokument, havner det lenger ned i resultatlisten enn i et mindre dokument der det gjentas flere ganger
  - Relevansalgoritmer kan godt kombinere dokumentkildene med katalogkilder, og f.eks. rangere en avlevering fra Gloppen kommune høyt hvis man søker på ordkombinasjonen *Ole + spesialundervisning + Gloppen*

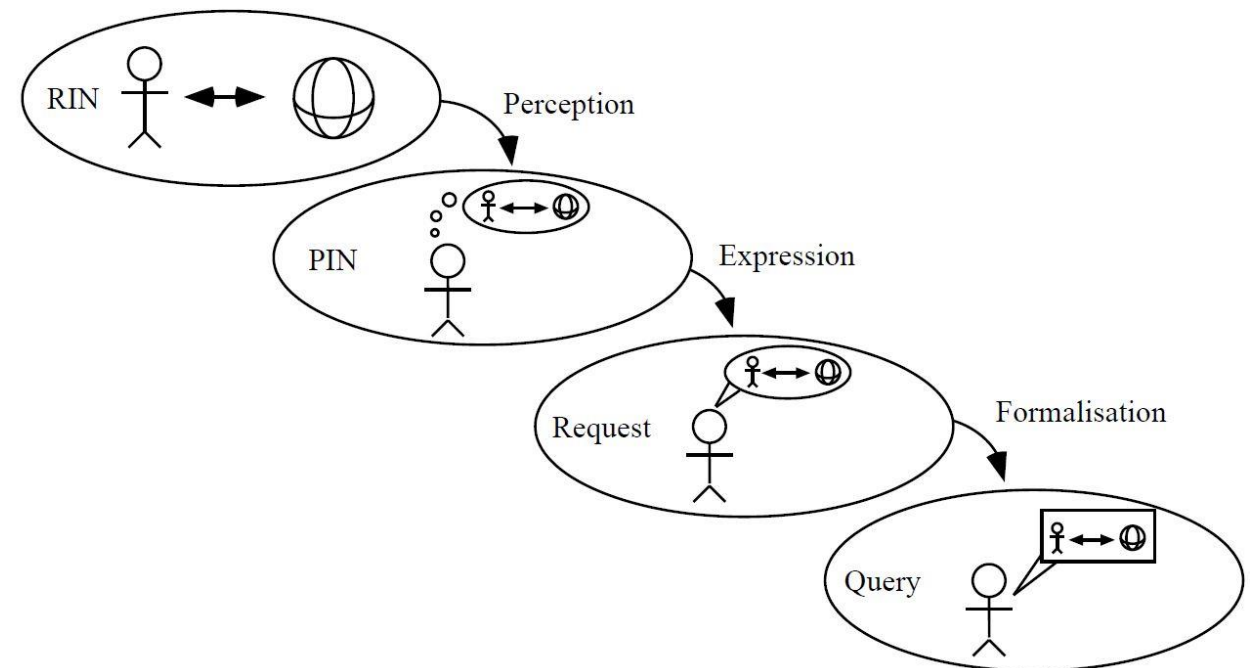
# Relevans innen IR er et eget forskningsfelt

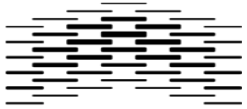
## – kanskje foreløpig litt fremmedartet i vårt fag

### Relevanstyper (Saracevic 2007)

Affektiv relevans	Det brukeren egentlig vil med informasjonssøket
Situasjonsrelevans	Hvilen situasjon/oppgave søket er en del av
Kognitiv relevans	Hva er det brukeren allerede vet?
Emne-relevans	Forholdet mellom spørsmålet og dokumentets emne
Algoritmisk relevans	Hvilke dokumenter systemet «tror» er relevant for søket

### Relevansens fire dimensjoner (Mizarro 1998)





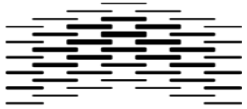
## Sammenligning mellom ulike fremfinningsstrategier

### Fremfinning via arkivstruktur

- Ordning og beskrivelser som er basert på faglige vurderinger
  - ekspertorganisert informasjon
- Kataloger og kompetanse/forkunnskaper kan gi innfallsvinkler til materiale som man ikke ville ha funnet ved å søke etter bestemte ord
- Avskjærer på en mer fundamental måte irrelevante eller utilgjengelige/klausulerte arkivenheter
- Kan i større grad gi belegg for å vurdere negativ troverdighet eller meningsfulle fravær
  - Struktur og kontekst som kilde til kunnskap

### IR, relevansrangerte treff i metadata og innhold

- Ikke-arkivkyndige brukere er mest vant til denne måten å finne frem i informasjon på
  - Kanskje arkivarer etter hvert også har blitt det?
- Mindre krav til brukerens forkunnskaper om aktører og funksjoner
  - Lettere å finne noe, men kanskje også vanskeligere å vurdere egen dømmekraft?
- Vesentlig mer automatiserbart og skalerbart enn å etablere katalogstrukturer
- Relevansbegrepet innen fulltekstsøk er mer brukerstyrt (subjektivt) enn katalogstrukturer



## Furner og Gililands «Archival IR» rammeverk (I)

- Information retrieval gir *intellektuell* tilgang til ressurser
  - Til forskjell fra fysisk og rettslig tilgang
  - IR-systemet avdekker hvilke ressurser som matcher det brukeren velger å søke etter
- Ressursbeskrivelser, autoritetsdata
  - Automatisk generert, faglig kontrollert, brukerbidrag
- Relevansrangering
  - Beregnet sannsynlighet for at noe er relevant
  - Lugger relevansbegrepet litt for arkivarer?
    - Furner og Gililand: «Selv om det kan være problematisk på et teoretisk/prinsipielt plan, er det uproblematisk for praktikere»



## Furner og Gililands «Archival IR» rammeverk (II)

- Postulat: En type «arkiv-information retrieval» kan og bør være en slags hybrid mellom hierarkisk struktur som inngang til materialet og et mer tekst- og relevansbasert tilslag på søketermer
  - Tettere kobling mellom beskrivelse og materiale, f.eks. med EAD-skjemaer (XML-skjema for arkivbeskrivelser, XML påtvinger alt hierarkiske strukturer)
- En slik hybrid form gir hierarkisk organiserte søkeresultater
  - Mindre brukervennlig enn relevansbasert IR (?)
  - Bedre muligheter for å vurdere hva man faktisk finner opp mot hva man mener man kunne ha ventet å finne, enn ved relevansbasert IR

## Kan arkivbeskrivelsene reddes?

- eller blir de irrelevante, eller økonomisk uforsvarlige, dersom IR gir resultater de fleste er fornøyde med?
- «Noe i toppen»-delen av beskrivelsen vil trenge som et *metaarkiv*
  - Man vil trenge å vite hvor arkivpakkene kom fra, av bevaringshensyn
- Hva vil være den mest nyttige bestanddelen av arkivbeskrivelser, når de kan brukes i kombinasjon med ulike former for innholdssøk?
  - Blir beskrivelsene bedre/mer relevante av å dra inn mer detaljinformasjon som speiler dokumentinnholdet? (EARK-prosjektets DIP-spesifikasjon)
  - Eller bør de først og fremst bidra med alternativ kunnskap om form/dokumenttyper? (Eddy Put)
  - Eller om funksjoner og virksomheter? (ICA R-i-C, Bearman, MacNeil)



## Oppsummering (I)

- Gjenfinning basert på arkivbeskrivelser, IR eller hybridvarianter
  - Alle tre posisjoner har sine talsmenn blant arkivteoretikere
  - Umulig å vite hvordan dette blir om noen år
    - Noe som vil vinne frem internasjonalt, eller ulik praksis i ulike land?
  - Jeg tror det vil haste å få opp kvaliteten på beskrivelsene av digitalt skapt materiale her til lands, hvis det skal ha en sjanse til å vinne frem som fremfinningsmetode

## Oppsummering (II)

- Å undervise i bruk og tilgjengeliggjøring av digitalt skapt materiale kan nærmest virke litt prematurt
  - Man kan sette sammen en del ulike elementer av kunnskap og ferdigheter, men resultatet blir ikke like praksisnært som andre emner
  - Lett å finne frem til dilemmaer, mulige stridsspørsmål og uløste problemer
- Så hvorfor gjør vi det?
  - Det trengs flere folk som er rustet til å være faglig bindeledd mellom arkivmaterialet og teknologien
  - Mer kunnskap om, og erfaring med, bruken av digitalt skapt materiale er også en forutsetning for å videreutvikle kunnskaper om danning og bevaring