

10. MAI 2017

# Uttrekk av historiske data

ESA OG EPHORTE

RAGNAR STURTZEL



# Agenda

- Prosjekt
- Datakvalitet
- Samspill
- Selve avleveringen

# Prosjekt

- Avleveringen er et prosjekt
  - Prosjektplan
  - Hvem skal involveres
  - Hva må gjøres i forkant
  - Avtaler med bl.a. depot
  - Evt datakvalitetskorrigeringer
  - Uttrekk
  - Overlevering
  - Oppfølging

# Datakvalitet

- Datakonsistens (teknisk)
  - Finnes refererte data (fortsatt)
  - Eksisterer alle brukte koder (fortsatt)
- Arkivmessig «konsistens»
  - Stemmer papir/elektronisk
  - Er det mulig å periodisere – bruk av arkivdeler
  - Finnes alle elektroniske dokumenter i arkivformat
  - Er alle restanser håndtert
  - Er alle internt produserte journalposter ekspedert
  - Er alle journalposter avsluttet
  - Er alle saker avsluttet

# Datakvalitet, erfaring 1

## Datakonsistens (teknisk)

- Kryssreferanser
  - kan være til tidligere avlevert materiale som så er slettet
  - kan være til materiale som skal avleveres senere
  - kan være til materiale som på annen måte er flyttet
- Opprinnelse/vedlikehold under andre regler
  - Data mangler
    - Var ikke en gang påtenkt den gang
  - Data (f.eks. sluttede brukere, koder som er gått ut av bruk) er slettet

# Datakvalitet, erfaring 2

## Arkivmessig «konsistens»

- Arkivet håndtert elektronisk, men er formelt på papir
- Arkivdeler stemmer ikke
- Mange «rare» dokumentformater
  - PDF/A-krav er «nye»
  - Feiloppsett ved konverteringer
  - Kan alt konverteres?
- Lange restanselister
  - Avskrive med SA?
- Er alle internt produserte journalposter ekspedert
  - Sikker på at journalposter i R har blitt sendt?
- Mange saker / journalposter er ikke avsluttet
  - Tvangsavslutte?

# Datakvalitet, erfaring 3

Kanskje kjøre en jevnlig helsesjekk på arkivet?

- Oppsett
- Rutiner
- Konsistens i referanser
- Status på saker, journalposter og dokumenter
- Restanser
- Arkivformat av elektroniske dokumenter

Lurt at arkivet blir sett på med «friske øyne»

# Datakvalitet, erfaring 4

Når rapportene viser avvik:

Hvilken kvalitet er bra nok?

- Tolkbart hos depotet?
- Vet man at en journalpost i R er ekspedert?
  - Hvorfor scripture til A da?
- 85 – 95 % av alt arkivverdig materiale er ikke i arkivet
  - Hvorfor bruke mye ressurser på «finpuss» av de 5 – 15 % som finnes?

Korrekt umiddelbart er bedre enn etterarbeid ved avlevering!



# Samspill 1

- Hvem skal involveres
- Hvordan skal de involveres
- Når skal de involveres
  
- Aktuelle
  - Arkivet
  - Arkivdepotet
  - Leverandør
  - Avleveringsspesialist

# Samspill 2

- Når?
  - Ved oppstart!
- Arkivet
  - Det er de som kjenner dataene og organisasjonen
- Arkivdepotet
  - Mye ekstraarbeid spart ved tidlig involvering
- Leverandør
  - Kjenner systemet og har gjort det før
- Ekstern spesialist
  - Kan være et alternativ
  - Oppdatert på løsningen?

# Avleveringen

- Hva skal avleveres
  - ESA kan ta med «alt»
  - ephorte tar kun avsluttede saker
- Uttrekk
- Rapporter
  - Fort gjort å glemme
- Testing
  - Testverktøy underveis
- Avviksrapport
  - Ref. helsesjekken
- Overlevering
- Tilbakemelding
  - Ofte stor jobb å tolke

# Uttrekk og rapporter

- ESA
  - Programmet K2Avlever
  - Definerer utplukk
  - Lager / starter avlevering
    - Noark 3
    - Noark 4.1
    - Noark 5.3.1
  - Lager / starter rapporter
- ephorte
  - Programmet Submission
  - Konfigurere
  - Kjøres steg for steg
  - Noark 5.3.1 (med siste versjon)

# Erfaring

- Inkonsistens i standardene
- Data er ikke perfekte
- Feil i uttrekksmodulene
- Testprogrammene er heller ikke perfekte!

# Standardene

- Ikke fullt ut samsvar mellom datatyper i xsd og i standarden
  - Gjelder spesielt datoformat
- Standarden åpner for egendefinerte verdier
  - xsd-ene krever at kun faste verdier benyttes
- Merk at Arkivverket ikke er ferdig med kravene til Noark 5.4 avlevering!

# Data

- Påkrevde data mangler (fortsatt)
  - Manglende tester under produksjon
  - Endret standard etter produksjon
  - Ikke fanget i helsesjekk
- Inkonsistenser
  - Gjelder helst Noark 4
  - Slettede data som ikke kunne / ble rekonstruert
- Ulovlige dokumentformater
  - Forskjellige verktøy gir forskjellige svar på konformitet
  - Feil i oppsett i verktøy
  - Feil i verktøy
  - Formater som ikke / vanskelig lar seg konvertere
- Men: Må data være 100 % korrekt?

# Testverktøy

- Arkn4
  - Krever at alle brukte postnummer ligger i postnummerregisteret
  - Krever at alle kortkoder for adresser (fortsatt) ligger i adresseregisteret
  - Krever at alle tilgangskoder finnes med hjemmel i hjemmelsregisteret (selv om de ikke brukes lenger)
- Arkade 5
  - Fortsatt under utvikling...
- Documaster
  - Svært dogmatisk
  - Fortsatt under utvikling...
- Teste selv? Teste tidlig?



# Uttrekk

- Stadige endringer i Noark
- Regelverket endret med «tilbakevirkende» kraft
- Data er svært varierende
- Gir større eller mindre avvik
- Men data er der!

# Oppsummering

- Stort prosjekt
  - Starter lenge før man tenker
  - Involver alle parter tidlig
  - Datakvalitet?
  - Så avlevering
- 
- Hvor mye ressurser skal brukes på prosessen



**EVRY**

Digital  
+ Advantage