

Del 4: Noark 5-uttrekks egnethet for langtidslagring

I den følgende teksten vil jeg drøfte påstanden om at et gyldig Noark 5-uttrekk er godt egnet for langtidslagring fordi det gjenspeiler arkivdanningen som fant sted, samtidig som det inneholder kontekst, struktur, innhold og stabilitet. To spørsmål kan utledes fra dette utgangspunktet: vil det virkelig være slik at et gyldig Noark 5-uttrekk gjenspeiler den faktiske arkivdanningen, og inneholder uttrekket tilstrekkelige og/eller de riktige metadataene for at uttrekket vil kunne tjene sitt formål i fremtiden?

Etter noe jeg antar var en ytterst uhøytidelig undersøkelse, skrev James Lappin (@JamesLappin), britisk blogger og *records manager*, følgende på Twitter 26. november 2013:

"It is interesting to compare the acquisition policies of NSA/GCHQ with those of NARA/TNA" – "the intelligence agencies are more interested in e-mail accounts than the formal record systems that provide the pipeline to the archives" – "I suspect historians would rather look at the e-mails than the formal record systems too"

Problemstillingen er relevant nok – det vi velger å bevare i de "formelle" arkivsystemene, er det som vil bli bevart for fremtidige generasjoner i depotinstitusjonene. Dette er, skal vi tro James Lappin, ikke tilstrekkelig for etterretningsvirksomhet – men er det likevel tilstrekkelig for fremtidens arkivbrukere? Bevarings- og kassasjonsspørsmål – dog ikke direkte tema for denne teksten – er blant de vanskeligste spørsmålene en arkivar kan forsøke å besvare. Vurderingene av hva som skal bevares og hva som ikke er verdig status som arkivmateriale, bør være grundige, faglig begrunnede, og kun gjennomføres av skikket personell – altså arkivarer. Likevel er det blitt slik at disse vurderingene er delegert til hver enkelt av oss – ikke arkivarene, men til hvert enkelt individ, hver enkelt ansatt med en e-postkonto eller en mobiltelefon. Det skjer umulig uten konsekvenser.

Mye har forandret seg de siste 100 årene, også dokumentasjons- og arkivpraksis. Der man tidligere skrev formelle brev, avholdt møter og ringte med telefonen (altså samtalte med stemmene våre), sender man i dag skriftlige, elektroniske meldinger frem og tilbake i noe som likner muntlige dialoger – kanskje skriver vi til og med "Hehe!" hvis noe har fått oss

til å le, og smilefjes med sur eller blid parentesmunn for å vise vår reaksjon på eller innstilling til den andres melding til oss. Vi ville aldri gjort noe tilsvarende i et formelt brev, selv om brevvekslingen skjer mellom saksbehandlere som ansikt til ansikt eller over telefonen vil kunne ha en vennskapelig tone. Forskjellen er at kommunikasjonen er skriftlig – "samtalene" dokumenteres, og blir dermed – når de oppstår som ledd i en persons gjerning som ansatt i offentlig virksomhet – gjenstand for både arkiv- og innsynsplikt. Det stopper likevel ikke der – samtalene dokumenteres i systemer som gir oss mindre pålitelig, gjenfinnbar og tilgjengelig arkivdokumentasjon enn noensinne tidligere fordi systemene som brukes er lukkede og proprietære, ikke bygget på anerkjente arkivfaglige prinsipper, og laget uten funksjoner for å skille arkivdokumentasjon fra annen informasjon og hindre manipulasjon eller andre uautoriserte endringer (Duranti 2010 s. 79). Den digitale kommunikasjonen kan dermed lett bli "[...] a continuous discourse without trace, as both act and record occur simultaneously with little or no media delay or survival." (Hugh Taylor, gjengitt i Duranti 1998 s. 70). Den største utfordringen med bruk av digitale arkivdannings- og kommunikasjonssystemer, er å klare å skape pålitelige arkivdokumenter, og deretter å bevare både arkivdokumentene og deres pålitelighet over tid (Duranti 2010 s. 80).

I norsk arkivtradisjon er det den formelle arkivdanningen og dennes kvalitet som avgjør hvilket arkivmateriale kommende generasjoner vil kunne benytte seg av (Riksarkivet 2013a s. 16). Arkivlovgivningen bestemmer rammene for arkivdanningen i offentlig virksomhet i Norge. Norsk arkivsystem/Norsk arkivstandard (Noark) bygger på dette lovverket, og stiller detaljerte krav til hvordan systemene som brukes i arkivdanningen skal legge til rette for at informasjonen og dataene som lagres skal kunne trekkes ut i teknologiavhengig format for langtidsbevaring. Uttrekket består av de faktiske arkivdokumentene i godkjent arkivformat, f.eks. PDF/A (så fremt arkivskaper ikke har hatt papirarkiv og kun ført journal elektronisk), samt metadatainformasjon om dem, i tillegg til journalrapporter og endringslogger. Både arkivdokumenter på papir og digitale arkivdokumenter består av elementene innhold, struktur, kontekst og presentasjon (Riksarkivet 2013a s. 22). Innholdet finnes i en eller flere filer som holder på selve "budskapet" og disse skal oppbevares på måter som gjør at fremtidige brukere kan forstå både budskapet og sammenhengen budskapet inngår i. Det er metadataene om innholdet, konteksten og strukturen som gjør dette mulig (Riksarkivet 2013a s. 22). Presentasjonen av informasjonen vil imidlertid endre seg avhengig av hvilken programvare som brukes

for å hente frem elementene som utgjør arkivdokumentet. Det er ikke uvanlig i dag at arkivdokumenter i samme sak eksisterer delvis på papir, delvis på e-post, delvis i et regneark og delvis i en relasjonsdatabase (Duranti 2010 s. 79). Å opprette og deretter bevare lenkene mellom de ulike elementene er vanskelig, men samtidig av avgjørende betydning for dokumentasjonen og informasjonsinnholdets verdi, både på kort og lang sikt. Når innholdet består av mer eller mindre implisitte elementer, kompliseres dette ytterligere. Rettsavgjørelser i USA har avvist arkivdokumentasjon i konvertert format fordi kritiske deler av dokumentasjonen kan bli borte i konverteringsprosessen – f.eks. vil ikke en papirutskrift eller en PDF/A-fil av et regneark inneholde formlene som beregningene i arket baserer seg på (Duranti 2010 s. 80). Dataenes pålitelighet og gyldighet skal kunne etterprøves (Øvrebø 2013), hvilket er umulig hvis kritisk informasjon har forsvunnet underveis i bevaringsprosessen. Spørsmålet blir dermed om Noark 5-uttrekket, slik standarden spesifiserer at det skal være, og for øvrig også de Noark-godkjente systemene som uttrekkene produseres fra, i tilstrekkelig grad bevarer informasjonsinnholdet i arkivdokumentene, slik dette informasjonsinnholdet blir til underveis i arkivdanningen. Derneft er spørsmålet om man evner å være seg denne problematikken bevisst, slik at man ved å stille krav til arkivdanningen, altså til måten dokumentasjonen produseres og lagres på, kan løse disse problemene på tilfredsstillende måter. For regneark kan arkivplanen f.eks. bestemme at beregninger gjengis og lagres eksplisitt i separate, tilknyttede dokumenter i arkivsystemet. Alternativet er at de går tapt.

Arkivdokumenter på papir har sin kontekstinformasjon implisitt i arkivsystemet, som oftest gjennom dokumentenes fysiske plassering sammen med andre dokumenter i samme sak eller i samme serie (Duranti 2009 s.45-46, fotnote 27). For digitale arkivdokumenter må informasjon om sammenhengen dokumentene inngår i, så vel som annen prosess- og strukturinformasjon, gjøres eksplisitt og kobles til dokumentets innhold og manifestasjon (Duranti 2009 s. 45-46). Forståelsen av informasjonsinnholdet – budskapet – avhenger av den foreliggende informasjonen om kontekst og struktur. F.eks. vil klassifikasjonsinformasjon og saks-/dokumentnummerering knytte dokumentet til andre dokumenter i samme sak eller serie, og informasjon om tidspunkter for mottak, ferdigstillelse og utsendelse vil plassere dokumentene i korrekt rekkefølge. Forutsatt at arkivdanningen er tilstrekkelig kontrollert og regelfestet, vil bruk av et Noark-godkjent system kunne sørge for at denne informasjonen blir lagret i tilknytning til informasjonsinnholdet, både mens systemet er i bruk og etter at dataene er trukket ut av

systemet og overført til depot. At dataene som blir lagret er sannferdige, er likevel ikke gitt. En digital enhet har et stabilt innhold hvis innholdet er uforandret og uforanderlig, altså at dataene ikke kan bli overskrevet, endret eller slettet (Duranti 2009 s. 47), og denne stabiliteten er avgjørende for arkivdokumentets verdi som bevis, fordi det er stabiliteten som bringer informasjonen – beviset – fremover i tid (Anderson 2012 s. 361). Ifølge Noark 5 skal dokumentenes pålitelighet sikres underveis i arkivdanningen av systemenes sporings- og loggfunksjoner, samt kravsatte grenser for hvilke innholdsendringer og andre endringer som skal kunne gjennomføres i de ulike fasene av dokumentproduksjonen, og deretter, som del av uttrekksproduksjonen, med hashfunksjoner som regner ut sjekksummer for samtlige dokumenter i uttrekket basert på innholdet i hvert dokument (Riksarkivet 2013 s. 250). Med unntak av enkelte minstekrav, legger standarden opp til at hver arkivskaper selv skal bestemme hvordan sporing og logging av hendelser i systemet skal skje underveis i arkivdanningen, mens uttrekks uforanderlighet sikres gjennom oppbyggingen av uttrekket og bruken av sjekksummer. Filen info.xml, som følger med et uttrekk som deponeres eller avleveres, skal inneholde overordnet informasjon om uttrekket slik arkivlovverket spesifiserer (Riksarkivet 2013 s. 93), blant annet en sjekksum for filen arkivuttrekk.xml, som i sin tur inneholder sjekksummer for alle metadatafilene i uttrekket. Filen arkivstruktur.xml skal, i tillegg til øvrige metadata, inneholde sjekksummer for alle arkivdokumentene i pakken. Når filen info.xml bevares separat fra resten av uttrekket, og tilgang til dette dokumentet er underlagt tilstrekkelig styring og kontroll, vil man, i alle fall i teorien, kunne sannsynliggjøre at dokumentene er uforandret innenfor et gitt tidsrom, og dermed er det de gir seg ut for å være. Verifikasjon av arkivmaterialets uforanderlighet og uforandrethet krever imidlertid mer enn strukturering av filene i uttrekket. Flere faktorer spiller inn og virker i kombinasjon; kontroll av systemimplementasjonen så vel som uttrekks konformitet med standardens føringer, generell og spesiell loggføring, sporbarhet og tilgangsstyring, og ikke minst nøkkelpersoners vitnesbyrd om at handlingene som har ledet frem til den manifestasjonen av informasjonen som man ønsker å feste lit til, har foregått i tråd med vedtatte prosedyrer (InterPARES Authenticity Task Force Report s. 22).

Noark 5 forutsetter, som beskrevet over, at arkivdokumentenes autentisitet sikres ved at dokumentene identifiseres, tas inn i systemet, tilknyttes metadata og deretter fryses for å hindre uautoriserte endringer (Riksarkivet 2013 s. 16). Denne prosessen er imidlertid ikke

unik for Noark. Andre standarder stiller tilsvarende krav, og andre tekniske løsninger har funksjonalitet som skal sikre arkivdokumentenes autentisitet over tid. Og det er ikke uten grunn. Denne "frysingsprosessen" regnes som noe av det viktigste som skjer i arkivdanningen, og det er denne prosessen som sørger for at arkivmaterialet er verdig den tilliten man gir det, og den verdien det har som informasjon, bevis eller begge deler, både på kort og lang sikt.¹ Dette kan nok virke temmelig rett frem hvis man studerer standarden og lovverket alene – vi kan ha tillit til at arkivdokumentet er det det gir seg ut for å være så lenge det er lagret i det Noark-godkjente systemet – men så enkelt er det ikke. Antakelsen man har om et arkivert dokumentets autentisitet baseres på et gitt eller fastsatt referansepunkt, f.eks. at dokumenter lagret i arkivsystemet er beskyttet mot endringer fordi systemets konfigurasjon hindrer slike endringer, mens den faktiske autentisiteten må, som antydning over, vurderes ut fra flere faktorer, også hva som har skjedd med dokumentet og budskapet det formidler på vei inn i arkivsystemet (InterPARES Authenticity Task Force Report s. 22). Arkivskapers øvrige systemer, også de ikke-tekniske, spiller inn her. I offentlig (arkivlovunderlagt) virksomhet i Norge er arkivplanen sentral, og kan være et viktig dokument også i autentisitetssammenheng, forutsatt at den inneholder retningslinjer for arkivdanningen, retningslinjene beskriver hvordan virksomheten forstår og tolker det juridiske miljøet den opererer innenfor, og også hvordan denne forståelsen/tolkningen skal operasjonaliseres i hvert enkelt ledd i arkivdanningsprosessen. Det er ikke tilstrekkelig at det elektroniske arkivsystemet hindrer brukere i å endre dokumentenes innhold uten at dette blir logget, hvis man ikke samtidig har kontrollerbare, eksplisitte tekniske eller andre rammer som bestemmer hvordan dokumentene skal komme til arkivsystemet. I hvilken grad et Noark 5-uttrekk inneholder tilstrekkelige metadata om kontekst, struktur, innhold og stabilitet er dermed nokså irrelevant, fordi informasjonen som metadataene er tilknyttet i seg selv er upålitelig. Arkivskaper lagrer arkivdokumenter i arkivsystemene med brutt *chain of custody* når

¹ Frysingsprosessen og dens betydning er omdiskutert i arkivteorien. Når en slik frysing av informasjon settes som vilkår for og mål på *recordness* (dvs. at bestemte egenskaper ved objektet underbygger dets verdi som arkivmateriale), ekskluderes samtidig alt som ligger utenfor det det er mulig å fange og fryse til fast form: "Dynamic non-fixed records are translated to a fixed form so that they can be recognized as having recordness. Instead, archivists should begin to recognize the forms of records specific to the community with which they wish to engage." (Anderson 2012 s. 361). Her omfattes muntlige kulturuttrykk som sang og dans, og disse uttrykkenes *recordness*, men man kan også ane konturene av en mulig ny tilnærming til "sosiale" kommunikasjonsformers anseelse som arkivmateriale.

dokumentene først har kommet til eller blitt produsert i, og deretter har eksistert i en ansatts e-postboks før det legges inn i arkivsystemet. Det foreligger ingen kontroll av hvorvidt dokumentet som lagres har det samme innholdet som dokumentet som ble mottatt eller sendt. Mulighetene for en slik inngripen fra arbeidsgivers side er svært begrensede, og kan vanskelig hjemles i lov så lenge arbeidsgiver ikke har mistanke om at det foreligger uregelmessigheter. Den ansattes registrering av dokumentet i arkivsystemet må dermed forstås som en autentisering av dokumentets innholds sannferdighet på registreringstidspunktet, uten at det i lovverket eller andre steder synes å oppleves nødvendig å stille spørsmål tegn ved dette vitnesbyrdet, eller etablere kontrollmekanismer som verifiserer den ansattes autentisering. Kanskje er det nettopp dette f.eks. NSA, i sin søken etter informasjon, ser ut til å ha forstått, når de regner e-postkontoene som mer relevante, og kanskje også mer autentiske, fullkomne og pålitelige, enn de formelle arkivsystemene. Gjengivelsen av blogger og *records manager* James Lappins twittermeldingseanse 26. november fullbyrdes:

"maybe we should allow NSA and GCHQ to carry on with their data-gathering on condition they also act as our national digital archives".

Kilder

Anderson, K. (2012). The footprint and the stepping foot: archival records, evidence, and time. *Archival Science*(13), 349-371.

Duranti, L. (2010). Concepts and principles for the management of electronic records, or records management theory is archival diplomatics. *Records Management Journal* 2010(1), 78-95.

Duranti, L. (2009). From Digital Diplomatics to Digital Records Forensics. *Archivaria* 68 (Fall 2009), 39-66.

Duranti, L. (1998). *Diplomatics. New uses for an old science*. USA, Maryland: Scarecrow Press, Inc.

InterPARES Authenticity Task Force Report. (udatert). Tilgjengelig 1. desember 2013 via interpares.org

jameslappin. (2013, 26. november). Twittermeldinger (gjengitt i teksten). Tilgjengelig 2. desember 2013 via <https://twitter.com/JamesLappin>

Riksarkivet. (2013). Noark 5. Standard for elektronisk arkiv. Versjon 3.1. (m. vedlegg). Tilgjengelig 3. november 2013 fra www.arkivverket.no

Øvrebø, O. A. (2013, 19. november). Sherlocks første lov om åpne data. I *Vox Publica*. Tilgjengelig 2. desember 2013 via http://voxpathlica.no/2013/11/sherlocks-foerste-lov-om-aapne-data/?utm_source=twitterfeed&utm_medium=twitter