



FAGARTIKKEL

Petter B. Høiaas, Cecilie Hansen Rørås og Thomas Sørdring

Uttrekkssammenligning – to uttrekk fra samme system

Migreringsstrategien er valgt for bevaring av digitale arkiver, og Norsk Arkivstandard (NOARK), definerer avleveringsstrukturen for avsluttede offentlige norske arkiver i dag. Denne standarden spesifiserer nødvendige metadata og struktur for et uttrekk. Gjeldende standard er Noark 5, men svært mye av det som kommer inn til depoter per dags dato skal avleveres etter Noark 4-standard. Testing av slike uttrekk gjøres med en rekke verktøy og arkivinstusjoner har til dels forskjellige praksis for hvordan man gjør disse testene. Felles for alle er at man tester den tekniske integriteten på selve uttrekket, men sammenligner det i liten grad med systemet uttrekket skal representere. Det gjøres lite eller ingen tester som etterprøver at man i uttrekket faktisk har fått med alle relevant data fra det originale systemet. Da Noark først og fremst spesifiserer et avleveringsformat, og databasene i sakarkivsystemene ikke nødvendigvis gjenspeiler denne strukturen, er det interessant å teste ulike uttrekk fra samme datagrunnlag opp mot hverandre.

Denne artikkelen fremstiller empiriske undersøkelser av to uttrekk tatt med forskjellige verktøy fra den samme databasen. Resultatene av sammenligningen gir et innblikk i hvordan uttrekksprosessen påvirker kvaliteten på dataene i arkivet. Med dagens testregime tester man først og fremst selve uttrekket, og det vil derfor være interessant å undersøke hvordan uttrekksmetode eller verktøy påvirker dataene i uttrekket.

Vi vil ikke i denne artikkelen gå inn på en diskusjon eller konsekvensanalyse av resultatene av denne sammenligningen. Dette vil kun være en gjennomgang av de empiriske undersøkelsene

som er foretatt på dette ene eksempelet. Diskusjoner rundt hva som er bakgrunnen for dette, hvilke konsekvenser det har og potensielt vil ha, samt hvordan dette kan unngås i fremtiden, vil være viktig å gjøre når man får mer informasjon enn det vi kan vise til etter test på ett system.

Bakgrunn og metode

Det ligger ingen større teoretisk eller metodisk avklaring som grunn for de empiriske undersøkelsene presentert her. Undersøkelsene er gjort ved IKA Kongsberg som ledd i testing av innlevert uttrekk av arkiv i henhold til Noark 4-standarden.

Gjennom et prosjekt ved IKA Kongsberg i samarbeid med Thomas Sødning, førsteamanuensis ved Høgskolen i Oslo og Akershus (HiOA), var det mulig å ta et alternativt uttrekk fra samme base som vi hadde mottatt uttrekk fra. Som bakgrunn for dette var både et ønske om å teste uttrekksverktøyet produsert i dette prosjektet, men også fordi det kunne vært interessant å se på hvilke ulikheter man kunne finne, om noen, i en slik sammenligning.

Både OAIS-standarden og TRAC-standarden anbefaler kvalitetskontroll ved mottak av et uttrekk ved depot, men de går ikke konkret inn på hva dette skal innebære. Slik testregimet fungerer i dag vil kvalitetskontrollene innebære blant annet virussjekk, integritetskontroll med sjekksummer, kontroll av filformater og om uttrekket er teknisk fullstendig og korrekt ved hjelp av for eksempel verktøyet ArkN4. Er testregimet nok for at uttrekket kan betegnes som fullstendig og korrekt? Hva har datakvaliteten på uttrekket å si for at et uttrekk skal kunne godkjennes som fullstendig og korrekt? Datakvalitet angir i hvilken grad data i et system er i overensstemmelse med det virkelige scenarioet dataen representerer. Dårlig datakvalitet kan oppstå om et systemet ikke er modellert slik at det representerer det virkelige scenarioet, eller det kan oppstå når prosedyrene for bruk av systemet i arkivdanningen er for dårlig. Depotet har ikke mulighet til å rette opp eventuelle feil eller mangler, og om det har gått for lang tid, vil det være for sent å gjøre noe med kvaliteten. Om for eksempel arkivdokumentets proveniens, kontekst og struktur er dårlig beskrevet, kan gjenfinning bli en vanskelig eller umulig oppgave. På bakgrunn av dette var det svært interessant å få mulighet til å se to uttrekk, produsert med ulike verktøy, av det samme datagrunnlaget opp mot hverandre. I prinsippet skulle de to uttrekkene være like, da databasen er den samme, og begge uttrekkene er strukturert i henhold til Noark 4-standarden.

Uttrekk A ble mottatt som normal deponering fra en av IKA Kongsbergs eierkommuner, som i samarbeid med flere andre kommuner skulle avslutte eldre systemer og oppgradere til Noark5. Dette uttrekket ble gjort av en kommersiell uttrekksleverandør. Da vi mottok uttrekk A, ba vi om en kopi av den originale basen fra kommunen, slik at vi hadde det originale datagrunnlaget som referanse. Fra denne basen ble det foretatt et uttrekk med uttrekksverktøyet produsert av Thomas Sødning.

På bestilling fra kommunen var det gjort enkelte opprydninger i basen før uttrekk A var foretatt. Behovet for dette var hovedsakelig grunnet en integrasjon med et fagsystem som hadde opprettet journalposter uten tilknytning til sak. I tillegg var det saker som ikke var blitt avsluttet og journaler som ikke var journalført på rett måte. På bestilling fra kommunen, og i samarbeid med

IKA Kongsberg, ble det derfor gjort opprydning av dette i basen før uttrekk A ble foretatt. Disse endringene ble gjort av uttrekksleverandøren på kopien oversendt for uttrekk A, og var derfor ikke gjort i datagrunnlaget for uttrekk B. Det var derfor ventet at det var enkelte inkonsekvenser mellom de to uttrekkene. Disse endringene var vel dokumenterte, og hvilke forskjeller som kunne forventes var derfor kjent.

Vi hadde tilgjengelig for test ett uttrekk produsert av en kommersiell uttrekksleverandør (uttrekk A) gjort fra en base etter enkelte endringer, én rå dump av den originale databasen satt opp i vårt test-miljø, og ett uttrekk gjort med verktøyet til Sødring (uttrekk B). Sammenligningene er gjort først og fremst på likheter og forskjeller mellom uttrekk A og uttrekk B, mens databasen fungerte som en referanse for data i begge uttrekk.

Sammenligningen baserer seg i all hovedsak på kvantifisering av tabeller og innslag i de to ulike uttrekkene. Dette er gjort i Notepad++ som har en enkel og effektiv funksjon for å finne antall innslag av attributt eller tekststreng i aktuelle fil. I tillegg er det gjort undersøkelser på ulikheter på enkeltinnføringer og tabeller i begge uttrekkene. Referansetester mot databasen er gjort ved SQL-spørringer i Oracle SQL Developer.

Overordnet analyse

En enkel sammenligning av hvilke xml-filer som er tatt med i uttrekk A og uttrekk B, viser 13 tabeller mer i uttrekk B enn i uttrekk A.¹ Alle de obligatoriske Noark 4-tabellene er inkludert i begge uttrekkene, basert på test med ArkN4. Nærmere undersøkelse viser at uttrekk B mangler 3 av tabellene som finnes i uttrekk A, hvilket gir et avvik på 16 manglende tabeller. Forskjellene ligger hovedsakelig i flere tabeller som knytter seg til utvalgsbehandling i Uttrekk B enn uttrekk A.

En sammenligning av antall attributter i like xml-filer i uttrekkene viser også ulikheter.² Denne typen sammenligning er ikke entydig, og det vil derfor være noe inkonsekvens å vente. Antallet attributter er ikke nødvendigvis lik fra innslag til innslag i den samme filen, og tallene vil derfor kun være en indikasjon på forskjeller. Det er lagt vekt på å se på antall attributter på samme innslag i de to uttrekkene så langt dette lar seg gjøre. Selv om forskjellene her for det meste er små er det enkelte filer som skiller seg ganske mye fra hverandre eksempelvis Avsender/mottaker-tabellen³ som har 11 attributter i uttrekk A og 6 i uttrekk B. Tabellene med antall attributter finnes i sin helhet i Tabell 2 i vedlegget.

Kvantifisering av innslag gir et bedre bilde av ulikheter mellom uttrekkene. De groveste tilfellene viser i overkant av 18 000 innslag mer i uttrekk B enn i uttrekk A.⁴ Uavhengig av innholdet i filene, viser det med all tydelighet at uttrekksprosessen, slik den er i dag, yter en betydelig

¹ Se vedlegg – tabell 1

² Se vedlegg – tabell 2

³ AVSMOT.XML

⁴ UTVBEHDO.XML

påvirkning på arkivet. Kvantifisering av innslag i hver xml-fil i de to uttrekkene finnes i Tabell 3 i vedlegget.

Kvantifisering av innslag alene avdekker ikke alle forskjeller mellom uttrekkene. Ved undersøkelse av arkivdelene i de to uttrekkene viser det seg at disse ikke er like.⁵ De har like mange innslag, men ett av innslagene skiller seg fra hverandre. Begge uttrekkene har oppdaget og adressert innslag i basen som ikke er knyttet til en arkivdel. Også journalenheter skiller seg fra hverandre, med et lignende avvik som i arkivdelene.⁶

Kvantitativ analyse – enkeltfiler

Blant de sammenligningene som kanskje klarest viser forskjellene mellom uttrekk A og B, er kvantifisering av innslag i xml-filene, og vi vil her presentere noen av de konkrete avvikene i litt større detalj.⁷

Av de kvantitative testene som viste størst avvik var telling av antall merknader. Uttrekk A viste 10 713 merknader, mens uttrekk B kun visste 255.⁸ Forskjellen har bakgrunn både i hvordan Noark 4 tabellen for merknader er bygd opp, og hvor informasjonen fra det aktuelle systemet må hentes. Noark 4 spesifiserer hvordan tabellen for merknader skal være bygd opp og hvilken informasjon som skal og kan inkluderes.⁹ Det aktuelle systemet har merknader fordelt på tre tabeller. Fra disse tabellene må uttrekksverktøyet hente data i ulike attributter, og strukturere dem i henhold til Noark 4-kravene.

Ved telling med SQL-kode av innslag fra de tre kjente tabellene som inneholder merknader, var det samlede antallet i basen 1588. Kvantifiseringen viste at uttrekk B sannsynligvis manglet den ene av de tre tabellene fra basen, noe vi også fikk bekreftet av produsenten av dette uttrekket. Det ble avklart at dette var som følge av at databasen uttrekksverktøyet var modellert over ikke hadde noen merknader i denne tabellen, og at den derfor var utelatt.

For uttrekk A var det mange fler merknader enn ventet. Tekstene var ikke typiske merknadstekster man skulle vente av saksbehandlings- eller arkivmerknader. I samtale med produsenten ble det avklart at i tillegg til de merknadene vi klarte å identifisere i basen, var det tatt med attributter som i utgangspunktet ikke var merknader. Valget var tatt på bakgrunn av erfaring med slike baser fra tidligere, som viste at informasjon kunne ligge i disse attributtene som kunne være viktig. Siden dette ikke hørte hjemme naturlig noe annet sted, ble det lagt til merknadene for å sikre at informasjon ikke forsvant.

⁵ ARKIVDEL.XML

⁶ JOURNENH.XML

⁷ For oversikt over kvantifisering av alle filer, se vedlegg – tabell 3

⁸ Tellingen er gjennomført i Notepad++. Antall er basert på antall åpnings-tager <MERKNAD>.

⁹ Noark 4 del 2, tekniske spesifikasjoner –MERKNAD.XML s.58 - 60. Link:

<http://arkivverket.no/arkivverket/Offentleg-forvaltning/Noark/Noark-4/Standarden>.

Ved siden av kvantitativ analyse av xml-filene, er det også gjort analyse av filer i større detalj. Flere filer viser inkonsekvenser i forhold til hverandre i måten data er strukturert og representert. Dette gjør sammenligning til tider krevende, men i mange tilfeller viser søk at man finner den samme data i begge uttrekk. Det er tidvis gjort forskjellige valg i de to uttrekkene som resulterer i hvordan xml-filene er sortert, men ved søk finner vi at det meste av data er tilgjengelig og rett strukturert i begge uttrekk. Noen analyser viser store forskjeller på data som ligger i de to uttrekkene. Under følger noen analyser av ulik data i like tabeller.

I merknadstabellen i uttrekk A var alle merknader taget som «arkivmerknad»¹⁰, mens det i uttrekk B kun var «saksbehandlermerknad»¹¹. At en base opererer med kun én slik type merknader er i seg selv oppsiktsvekkende, og at uttrekkene hadde forskjellig type-bemerkning på de samme merknadene kunne indikere at dette hadde en sammenheng med uttrekksprosessen. I samtale med uttrekksleverandør av uttrekk A kom vi frem til at det var feil at merknadene skulle være «arkivmerknader», basert på informasjonsdeklarasjonen i basen.¹² Spørsmålet var fortsatt hvordan det var ønskelig å merke de resterende merknadene. Ved gjennomgang av de andre informasjonskildene ble det tydelig at disse tabellene ikke hadde attributter som deklarerer merknadstypen.¹³ Siden Noark 4 krever en slik merknadstype er uttrekksleverandøren nødt til å ta et valg, og tillegge informasjon til uttrekket som ikke finnes i det originale materialet.

I adresse-filen¹⁴ i uttrekk A er alle adresser merket som utenlands-adresser.¹⁵ At kommunen skulle ha praksis å registrere adressene som utenlands-adresser virker lite sannsynlig. Sammenligning viste at adressene i uttrekk B er registrert med vanlig postnummer og poststed.¹⁶ Dette er en indikasjon på at det har blitt foretatt konvertering av data i uttrekksprosessen. I sammenligning med basen var det uttrekk B som gjenspeilet arkivskapers praksis best. På bakgrunn av dette ble uttrekksleverandør for uttrekk A kontaktet for å rapportere om feilen, og høre om dette var en tilsiktet endring. Det ble klart at uttrekksleverandør hadde oversett en endring i sitt verktøy som var gjort for en tidligere kunde. Dette førte til at alle adresser ble merket som utenlands-adresser.

Ved detaljert gjennomgang av sak- og journaltabellene¹⁷ i uttrekk A ble det klart at alle saker og journaler vi arbeidet med var merket som lagret på papir.¹⁸ Arkivet skulle være helelektronisk, og var levert med digitale dokumenter. Systemet var allikevel i bruk i en periode der det kunne være

¹⁰ INFOTYPE: «MA»

¹¹ INFYTYPE: «MS»

¹² Tabellen for kommentarer i basen viste tydelig at de 15 innslagene i tabellen var av typen MS - Saksbehandlermerknad

¹³ Gjennomgang av sakstabell og journaltabell i basen viser egen attributt «Merknad», men ingen «Itype».

¹⁴ ADRESSEK.XML

¹⁵ Lagt til attributt <AK.UTLAND> med prefiks NO eks: <AK.UTLAND>NO-0114 OSLO</AK.UTLAND>

¹⁶ <AK.POSTNR>0114</AK.POSTNR> og <AK.POSTSTED>OSLO</AK.POSTSTED>

¹⁷ NOARKSAK.XML og JOURNPOST.XML

¹⁸ Alle saker og journaler var merket med <SA.PAPIR>1</SA.PAPIR> og <JP.PAPIR>1</JP.PAPIR>

enkelte fysiske originaler som skulle være en del av arkivet. Vi gjorde derfor en kvantitativ analyse av antall innslag i sak- og journaltabellene¹⁹ som var markert som papirdokumenter.²⁰

Undersøkelsene viste at alle saker og alle journalinnslag i uttrekket var merket med «...*alle dokumentene tilknyttet denne saken skal arkiveres på papir,...*».²¹ Ved like analyser på uttrekk B, kunne vi konstatere at tallene var svært annerledes. Sammenligning med basen viste at informasjonen i uttrekk A var feil. Enkelte saker og journaler inneholdt referanser til papirdokumenter, men i et mye mindre omfang enn det kom frem i uttrekk A. I samtale med leverandør av uttrekk A, ble det klart at dette var en feil som hadde sin bakgrunn i at en tidligere kunde hadde ønsket en omskrivning av dette feltet. På bakgrunn av dette hadde denne omskrivingen også blitt gjort i dette uttrekket.

Dokumenter

ArkN4-rapportens punkt 8.1. Sammendrag konstaterte at «*Ingen dokumenter var nevnt i DOKVERSJON.*» Stikkprøver og manuelle tester ble gjennomført for å undersøke om integriteten mellom xml-filene og filene som var levert, var intakt. Dette ble funnet i orden, og antallet leverte dokumenter som var funnet av ArkN4 stemte overens med dokumentlageret i uttrekket, samt det som stod i informasjons-fil og ved telling av antall innslag i tabellen som skal inneholde koblingen mellom metadata og dokumenter.²²

En kvantitative analyse av dokumentversjon-tabellen satt opp mot uttrekk B, viste et betydelig avvik. Uttrekk A viste 28 667 filer, mens uttrekk B viste 44 477. Et avvik på 15 810 filer mellom de to uttrekkene. På bakgrunn av dette ble det utført kvantitativ analyse av innslagene i basen. Det ble gjort søk etter unike filid-er i dokumentlageret, samt av unike filid-er i e-postlageret. Begge disse inneholdt dokumentfiler, og til sammen viste de 57 857 unike filid-er i basen.

I avklaring med uttrekksleverandør av uttrekk A ble det klart at det kun var *aktive* filer som var prioritert i uttrekket. Det ble derfor i samarbeid med den aktuelle leverandør gjort en analyse av hvor mange aktive filer som lå i basen. Resultatet var 47 646 aktive filer, som derfor skulle vært med i uttrekket, altså et avvik på 18 979 filer.

Konklusjon

Til tross for svært enkle analyser av materialet, viser analysene presentert over at uttrekksprosessens påvirkning av arkivmaterialet er betydelig i dette tilfellet. Det resulterer blant annet i 18 979 manglende filer i uttrekket A. Dette er et avvik som heller ikke hadde blitt

¹⁹ NOARKSAK.XML og JOURNPST.XML

²⁰ Antall innslag av <SA.PAPIR>1</SA.PAPIR> og <JP.PAPIR>1</JP.PAPIR>

²¹ Noark 4 – del 2 tekniske spesifikasjon, s.12. Link: <http://arkivverket.no/arkivverket/Offentleg-forvalting/Noark/Noark-4/Standarden>.

²² DOKVERS.XML

oppdaget om det ikke var mulig å sammenligne uttrekket med et alternativt uttrekk eller den originale basen. Det var teoretisk mulig å godkjenne dette uttrekket på bakgrunn av gjeldende tester. Da dette uttrekket er første av flere samarbeidende kommuner, der de andre venter på behandling av sine uttrekk, ligger det et visst press på å godkjenne uttrekkene. Under like omstendigheter og uten mulighet til å kvalitetssikre uttrekk A mot uttrekk B eller basen, ville uttrekket mest sannsynlig blitt godkjent. Kommunens arkiver for en lang periode hadde derfor inneholdt mangler og feil som hadde gjort innsyn og gjenbruk svært vanskelig og til en viss grad umulig.

Undersøkelsene i denne artikkelen er svært begrensede, og de er gjort uten en grundig avklaring av teori og metode, men gir en klar indikasjon på at denne typen undersøkelser bør gjøres i større omfang. Når et uttrekk produseres gjøres det valg og endringer, både tilsiktede og utilsiktede, i arkivet. Det må foretas flere og større undersøkelser på denne typen datakvalitet på flere ulike systemer for å avdekke om dette er et enestående eksempel, eller om dette er gjennomgående for uttrekk som blir bevart i dag. For å sikre arkivene på best mulig måte må det foretas grundigere tester av de prosessene som har påvirkning på det uttrekket som blir gjort. Uttrekkprosessen kan sammenlignes med den tradisjonelle ordningen av et arkiv, der en arkivar rydder og begrenser arkivet. Med dagens praksis har vi overlatt arkivordningen til konsulenter med grundig forståelse av database og IT-teknologi, men uten realkompetanse på arkiv. Dette krever arkivfaglig og kvalitetsmessig testing av alle uttrekk av arkivarer med arkivfaglige kriterier. Med dagens testregime kan man enkelt si at arkivinstitusjonene hovedsakelig tester kvaliteten på «boksene» arkivet er lagret i, uten å gjøre tester på om innholdet i boksene er komplett eller riktig.

Vedlegg

Tabell 1 - Sammenstilling av inkluderte XML-filer i uttrekk A og B.

Uttrekk A	Uttrekk B	MERKNAD	MERKNAD
XML-filer	XML-filer	MOTDOKL	
ADMINDEL	ADMINDEL	NOARKSAK	NOARKSAK
	ADRADMEN	NUMSERIE	NUMSERIE
ADRESSEK	ADRESSEK	OPRITYPE	OPRITYPE
	ADRPERS	ORDNPRI	ORDNPRI
ADRTYPE	ADRTYPE	ORDNVERD	ORDNVERD
	ALIASADM	PERKLAR	
ARKIV	ARKIV	PERNAVN	PERNAVN
ARKIVDEL	ARKIVDEL	PERROLLE	PERROLLE
ARKIVPER	ARKIVPER	PERSON	PERSON
ARSTATUS	ARSTATUS	POLSAKG	POLSAKG
AVGRKODE	AVGRKODE		POSTNR
AVSKRM	AVSKRM		SAKPART
AVSMOT	AVSMOT	SAKSTAT	SAKSTAT
BSKODE	BSKODE	SAKTYPE	SAKTYPE
DOKBESK	DOKBESK		STATMDOK
DOKKAT	DOKKAT	TGGRP	TGGRP
DOKLINK	DOKLINK	TGHJEM	TGHJEM
DOKSTAT	DOKSTAT		TGINFO
DOKTILKN	DOKTILKN	TGKODE	TGKODE
DOKTYPE	DOKTYPE	TGMEDLEM	TGMEDLEM
DOKVERS	DOKVERS		TLKODE
	EARKKODE	UTDOKTYP	UTDOKTYP
	EMNEORD	UTVALG	UTVALG
ENHTYPE	ENHTYPE		UTVBEH
FORSMATE	FORSMATE	UTVBEHDO	UTVBEHDO
	FSTATUS		UTVBEHST
INFOTYPE	INFOTYPE	UTVMDOK	UTVMDOK
JENARKD	JENARKD	UTVMEDL	UTVMEDL
JOURNENH	JOURNENH		UTVMEDLF
JOURNPST	JOURNPST	UTVMOTE	UTVMOTE
JOURNSTA	JOURNSTA		UTVSAK
KASSKODE	KASSKODE	UTVSAKTY	UTVSAKTY
KLASS	KLASS	UTVTYPE	
	LAGRENH	VARAFORM	VARAFORM
LAGRFORM	LAGRFORM		
MEDADRGR	MEDADRGR		

56

69

Tabell 1 – Sammenstilling av hvilke XML-filer som er inkludert i de to uttrekkene. Antall filer er summert i bunn. Alle filer er stilt ved siden av sin like i motstående kolonne. Åpne felt indikerer derfor «manglende» XML-filer i forhold til det andre uttrekket

Tabell 2 - XML-filer med kvantifisering av attributter.

Uttrekk A		Uttrekk B	
XML-filer	Antall att.	XML-filer	Antall att.
ADMINDEL	7	ADMINDEL	5
		ADRADMEN	0
ADRESSEK	7	ADRESSEK	5
		ADRPERS	0
ADRTYPE	2	ADRTYPE	2
		ALIASADM	3
ARKIV	5	ARKIV	5
ARKIVDEL	10	ARKIVDEL	10
ARKIVPER	4	ARKIVPER	5
ARSTATUS	5	ARSTATUS	5
AVGRKODE	2	AVGRKODE	2
AVSKRM	4	AVSKRM	4
AVSMOT	11	AVSMOT	6
BSKODE	2	BSKODE	2
DOKBESK	6	DOKBESK	6
DOKKAT	2	DOKKAT	2
DOKLINK	6	DOKLINK	6
DOKSTAT	2	DOKSTAT	2
DOKTILKN	4	DOKTILKN	4
DOKTYPE	6	DOKTYPE	2
DOKVERS	8	DOKVERS	8
		EARKKODE	4
		EMNEORD	1
ENHTYPE	3	ENHTYPE	3
FORSMATE	2	FORSMATE	2
		FSTATUS	2
INFOTYPE	6	INFOTYPE	5
JENARKD	2	JENARKD	2
JOURNENH	6	JOURNENH	2
JOURNPST	17	JOURNPST	11
JOURNSTA	10	JOURNSTA	10
KASSKODE	2	KASSKODE	2
KLASS	5	KLASS	5
		LAGRENH	2
LAGRFORM	4	LAGRFORM	4
MEDADRGR	2	MEDADRGR	2
MERKNAD	9	MERKNAD	9
MOTDOKL	6		

NOARKSAK	13
NUMSERIE	6
OPRITYPE	2
ORDNPRI	11
ORDNVERD	5
PERKLAR	6
PERNAVN	5
PERROLLE	6
PERSON	3
POLSAKG	7
SAKSTAT	5
SAKTYPE	4
TGGRP	5
TGHJEM	3
TGKODE	4
TGMEDLEM	6
UTDOKTYP	2
UTVALG	10
UTVBEHDO	4
UTVMDOK	11
UTVMEDL	10
UTVMOTE	5
UTVSAKTY	2
UTVTYPE	2
VARAFORM	2

NOARKSAK	14
NUMSERIE	6
OPRITYPE	2
ORDNPRI	10
ORDNVERD	5
PERNAVN	7
PERROLLE	6
PERSON	2
POLSAKG	7
POSTNR	4
SAKPART	3
SAKSTAT	2
SAKTYPE	4
STATMDOK	2
TGGRP	6
TGHJEM	4
TGINFO	5
TGKODE	3
TGMEDLEM	3
TLKODE	2
UTDOKTYP	2
UTVALG	5
UTVBEH	10
UTVBEHDO	6
UTVBEHST	7
UTVMDOK	12
UTVMEDL	8
UTVMEDLF	6
UTVMOTE	5
UTVSAK	6
UTVSAKTY	2
VARAFORM	2

Tabell 2 - Tabellen viser alle inkluderte XML-filer i de to uttrekkene, samt kvantifisering av antall innslag og attributter i hver enkelt fil. Attributter kan variere fra innslag til innslag, og er derfor en lite korrekt måte å se forskjeller på dette nivået. Grønne felt indikerer at antallet er likt i begge uttrekk, røde indikerer forskjell.

Tabell 3 - Tabell med kvantifisering av antall innslag med differanse.

Uttrekk A		Uttrekk B		Avvik innslag
XML-filer	Antll innslag	XML-filer	Antall innslag	
ADMINDEL	42	ADMINDEL	42	0
		ADRADMEN	0	0
ADRESSEK	1457	ADRESSEK	1374	83
		ADRPERS	0	0
ADRTYPE	8	ADRTYPE	6	2
		ALIASADM	45	-45
ARKIV	1	ARKIV	2	-1
ARKIVDEL	13	ARKIVDEL	13	0
ARKIVPER	2	ARKIVPER	2	0
ARSTATUS	4	ARSTATUS	4	0
AVGRKODE	5	AVGRKODE	5	0
AVSKRM	8	AVSKRM	8	0
AVSMOT	55571	AVSMOT	42846	12725
BSKODE	2	BSKODE	2	0
DOKBESK	39034	DOKBESK	44462	-5428
DOKKAT	23	DOKKAT	15	8
DOKLINK	39034	DOKLINK	43736	-4702
DOKSTAT	2	DOKSTAT	2	0
DOKTILKN	18	DOKTILKN	8	10
DOKTYPE	12	DOKTYPE	12	0
DOKVERS	28667	DOKVERS	44477	-15810
		EARKKODE	4039	-4039
		EMNEORD	4039	-4039
ENHTYPE	4	ENHTYPE	3	1
FORSMATE	15	FORSMATE	6	9
		FSTATUS	5	-5
INFOTYPE	9	INFOTYPE	39	-30
JENARKD	11	JENARKD	10	1
JOURNENH	12	JOURNENH	12	0
JOURNPST	33230	JOURNPST	33369	-139
JOURNSTA	11	JOURNSTA	13	-2
KASSKODE	4	KASSKODE	4	0
KLASS	11819	KLASS	11956	-137
		LAGRENH	1	-1
LAGRFORM	13	LAGRFORM	6	7
MEDADRGR	1050	MEDADRGR	1036	14
MERKNAD	10713	MERKNAD	255	10458
MOTDOKL	1652			1652

NOARKSAK	8945	NOARKSAK	8935	10
NUMSERIE	102	NUMSERIE	1	101
OPRITYPE	8	OPRITYPE	8	0
ORDNPRI	10	ORDNPRI	9	1
ORDNVERD	3201	ORDNVERD	1695	1506
PERKLAR	540			540
PERNAVN	203	PERNAVN	202	1
PERROLLE	59	PERROLLE	201	-142
PERSON	203	PERSON	203	0
POLSAKG	2118	POLSAKG	1641	477
		POSTNR	4595	-4595
		SAKPART	7026	-7026
SAKSTAT	6	SAKSTAT	7	-1
SAKTYPE	10	SAKTYPE	16	-6
		STATMDOK	4	-4
TGGRP	43	TGGRP	95	-52
TGHJEM	62	TGHJEM	17	45
		TGINFO	246	-246
TGKODE	18	TGKODE	11	7
TGMEDLEM	207	TGMEDLEM	259	-52
		TLKODE	58	-58
UTDOKTYP	14	UTDOKTYP	11	3
UTVALG	31	UTVALG	32	-1
		UTVBEH	2402	-2402
UTVBEHDO	1615	UTVBEHDO	19742	-18127
		UTVBEHST	10	-10
UTVMDOK	1652	UTVMDOK	2466	-814
UTVMEDL	317	UTVMEDL	317	0
		UTVMEDLF	5	-5
UTVMOTE	568	UTVMOTE	770	-202
		UTVSAK	2398	-2398
UTVSAKTY	4	UTVSAKTY	6	-2
UTVTYPE	8			8
VARAFORM	5	VARAFORM	4	1

Tabell 3 – Sammenstilling av inkluderte XML-filer med utregnet innslagsavvik. Plussverdier indikerer flere innslag i uttrekk A enn uttrekk B. Minusverdier tilsier flere innslag i uttrekk B enn i uttrekk A.