

**En informationsvidenskabelig og arkivteoretisk
undersøgelse af arkivalier**

og

**forslag til dokumentation af data
i systemuafhængige arkiveringsversioner
af elektroniske arkivalier**

Lone Smith Jespersen
Januar 2005

UMTS-projekt
Digitalisering af dokumentation af IT-arkivalier

It is the task of the archivist to make a representation of the original context, so that researchers will be able not only to retrieve information, but above all to interpret the information in the original, administrative, functional context. That is what archivists are on this earth for, that is their added value.

(Peter Horsman i The Principle of Provenance)

INDHOLDSFORTEGNELSE

0. Indledning	4
1. Infologi – det teoretiske grundlag	7
1.1. Data, information og den infologiske ligning	7
1.2. Den infologiske model	9
1.3. Fortolkning og forudviden	12
2. Arkivalier fra en infologisk og en arkivistisk synsvinkel	14
2.1. Begrebet arkivalier	14
2.2. Data, information og arkivalier	16
2.3. Proveniensbegrebet og elektroniske arkivalier	19
2.4. Den infologiske model og elektroniske arkivalier	24
2.5. Autenticitet	28
2.6. Opsummering	30
3. Internationale projekter og anbefalinger	31
3.1. Open Archival Information System (OAIS)	31
3.2. InterPARES	33
3.3. International Council on Archives: Guiden 1997 og Håndbogen 2004	37
3.4. ISAD(G): General International Standard Archival Description	39
3.5. Dataarkiver	43
4. Forslag til den ideelle dokumentation	45
4.1. Begrebet dokumentation og formålene med dokumentationen	46
4.2. Oprindelig eller arkivskabt dokumentation	47
4.3. Den ideelle liste	49
5. Forslag til retningslinier for dokumentation i det praktiske arbejde med e-arkivalier	56
5.1. Oprindelig dokumentation (aktørskabt)	56
5.2. Arkivrettet dokumentation (aktørskabt)	57
5.3. Arkivskabt dokumentation	60
5.4. Placering af dokumentationen	61
5.5. Ressourcer	62
6. Forslag til digitalisering af dokumentation af ældre e-arkivalier	63
7. Afrunding	66
Litteraturliste	68

0. INDLEDNING

Denne rapport er resultatet af et halvt års en persons-projekt med den oprindelige titel Digitalisering af dokumentation af IT-arkivalier. Ifølge det oprindelige projektoplæg havde det to formål: For det første at komme med forslag til, hvorledes den papirbase-rede dokumentation af afleveringer af IT-arkivalier før regelsæt af 2000 kan digitalise-res. For det andet at komme med også fremadrettede forslag til indholdet, formen og strukturen på den såkaldte generelle information, det vil sige den dokumentation, der indgår i en arkiveringsversion som supplement til det systemuafhængige dataudtræk fra IT-systemet.

Der var oprindeligt lagt op til et samarbejde med projektet vedrørende påbegyndelse af format- og strukturkonvertering samt et forskningsprojekt med titlen Anvendelse af IT-arkivalier og konsekvenserne af et IT-perspektiv for historisk metode. Desværre blev forskningsprojektet ikke realiseret, og der blev ikke et tidssammenfald med påbegyndel-sen af format- og strukturkonvertering. Dette projekt er derfor desværre forløbet uden mulighed for samarbejde med beslægtede projekter. Dog har jeg haft kontakt med ERI-CA-projektet¹ og fået inspiration fra dette projekts case-studies i ældre registre.

Det fremgår af projektbeskrivelsen, at forslagene skulle udarbejdes på basis af erfaringer fra undersøgelse af konkrete e-arkivalier. Da jeg gik i gang med projektet, måtte jeg imidlertid erkende, at en sådan induktiv fremgangsmåde ikke ville være frugtbar. For det første er det problematisk at vurdere, hvilken dokumentation der er nødvendig for at anvende et e-arkivalie, når det ikke sker i forbindelse med et konkret forskningsprojekt, hvor e-arkivaliet indgår som datagrundlag eller kilde. For det andet kan det være pro-blematisk at afgøre, hvor mange og hvilke e-arkivalier der skal undersøges for at kunne generalisere erfaringerne.

I stedet har jeg valgt ud fra en teoretisk synsvinkel at diskutere følgende spørgsmål og ud fra svarene på dem at aflede, hvilke former for dokumentation der ideelt set bør led-sage systemuafhængige arkiveringsversioner:

- Nogle specifikke kendetegn ved arkivalier i forhold til anden information.
- Forskelle på arkivalier på papir og elektroniske arkivalier.
- Forskelle på IT-systemer i brug hos en arkivskaber og systemuafhængige arkive-ringsversioner af IT-systemerne.

Som teoretisk synsvinkel har jeg valgt begrebet infologi. Det er et ret ukendt begreb, der er opfundet af Börje Langefors, der helt tilbage fra 1950'erne har spillet en fremtræden-de rolle inden for systemudvikling i Sverige og resten af Skandinavien. Langefors var fra midten af 1960'erne og til omkring 1980 professor i informatik ved Kungliga tek-niska högskolan i Stockholm, og han tilskrives at være ophavsmand til det svenske ord for computer, 'dator'. Langefors har skrevet en lang række bøger og artikler om sy-

¹ Indførelse af elektroniske registre i centraladministrationen – en forvaltningshistorisk og arkivistisk undersøgelse.

stemudvikling ikke kun fra en datalogisk synsvinkel, men også ud fra synsvinklen om sammenhængen mellem teknik, opgaver, mennesker og organisation. I den forbindelse har han udviklet en teori, som han kalder infologi, der som centrale begreber har data, information, fortolkning, forudviden og kommunikationsfælleskab.

I første hovedafsnit vil jeg give en introduktion til disse begreber ud fra Langefors' værker. Langefors gør sjældent rede for, hvor han har sin inspiration fra, og han henviser kun i begrænset omfang til anden litteratur, men der er ingen tvivl om, at han har sin inspiration mange steder fra. Det har været fristende at grave dybere i de teoretiske forudsætninger samt at diskutere og problematisere begreberne ud fra erkendelsesteoretiske og sproglige teorier. Det har jeg imidlertid måttet afstå fra, da det ikke har været muligt inden for den givne tid, og da jeg har måttet holde mig for øje, at mit mål var at opstille konkrete og operationelle forslag.

I hovedafsnit 2 beskriver jeg arkivalier fra en infologisk synsvinkel for at besvare de tre spørgsmål ovenfor og for at få et grundlag for at definere, hvad der ideelt set skal dokumenteres ved e-arkivalier. Samtidig søger jeg at trække en forbindelse til arkivbegreber som arkivalie, proveniens og autenticitet, og jeg berører de diskussioner, der internationalt har været siden begyndelsen af 1990'erne om e-arkivalier.

Der har været mange diskussioner, som især er afspejlet i artikler og indlæg på konferencer, og når det gælder den tidligste litteratur, har jeg især fået et indtryk af den gennem den sammenfatning, som blev lavet i ICA-regi i 1996.² Jeg har naturligvis især ledt efter indlæg, der kunne give inspiration i forhold til de specifikke spørgsmål og problemstillinger i dette projekt. Noget har jeg fundet, men især de første år synes præget af en vis forvirring over e-arkivalier som nyt fænomen, og indlæggene peger i mange forskellige retninger, også fordi der undertiden ikke forstås det samme ved de samme begreber.

De senere år er der mere konkret arbejdet med projekter og retningslinier for bevaring af e-arkivalier. I hovedafsnit 3 redegør jeg for nogle af disse projekter og retningslinier, som jeg først og fremmest har beskæftiget mig med for at få inspiration. Det skal dog understreges, at jeg ikke har undersøgt tekniske retningslinier og standarder, da det ikke er emnet for dette projekt. Det skal også understreges, at der findes væsentlige projekter, som jeg ikke berører, da jeg også her har måttet begrænse mig.

I hovedafsnit 4 uddrager jeg de typer af dokumentation, som jeg undervejs i hovedafsnit 2 har peget på som væsentlige i forhold til e-arkivalier. Dette er at betragte som den ideelle liste, der som det meste ideelle i denne verden ikke kan realiseres. Derfor vurderer og diskuterer jeg i hovedafsnit 5 listen ud fra et Statens Arkiver her-og-nu synspunkt, ligesom jeg kommer med forslag til, hvordan den kan implementeres, hvem der skal dokumentere hvad, samt hvor og hvordan det kan gøres.

I hovedafsnit 6 kommer jeg med forslag til, hvordan den dokumentation, der ved tidligere arkiveringsversioner er afleveret på papir, kan digitaliseres.

² Erlandsson, Alf, *Electronic Records Management: A Literature Review*

Rapporten indeholder således afsnit af forskellig karakter: Hovedafsnit 1, 2 og 4 er den informationsteoretiske og arkivalske undersøgelse af fænomenet e-arkivalier samt de behov, der ideelt set er for dokumentation af data i systemuafhængige arkiveringsversioner. I hovedafsnit 5 og 6 er fokus på det konkrete og operationelle i forhold til Statens Arkiver, mens hovedafsnit 3 er min videreformidling af nogle af de mest vægtige projekter med relation til e-arkivalier.

1. INFOLOGI - DET TEORETISKE GRUNDLAG

Meget af det følgende vil kredse omkring det at forstå, hvad der menes med tegn og sammenstillinger af tegn. Det er også mit helt konkrete problem, når jeg sidder og skriver dette – hvordan sikrer jeg bedst, at den information, som jeg forsøger at formidle, faktisk bliver forstået, altså som jeg selv forstår den i mit hoved? Blot det at anvende et begreb kan volde problemer, bl.a. fordi begreber i tilknytning til elektroniske arkivalier kan tolkes forskelligt, afhængig af en datalogisk eller arkivistisk indgangsvinkel. Som eksempel kan nævnes begrebet record, som jeg senere vender tilbage til. I et forsøg på at undgå begrebsmisforståelser vil jeg i det følgende ud over definitioner i en del tilfælde anvende begreber, som ellers ikke bruges, i et forfængeligt håb om at jeg så i højere grad kan bestemme, hvordan læseren skal forstå dem!

Infologi er et eksempel på et sådant begreb, som stort set kun er blevet anvendt af Langefors, der sjældent definerer det. Det er dannet som en modsætning til datalogi, og et enkelt sted defineres det som ”vor viden eller vore kundskaber om, hvordan information tilvejebringes og udnyttes.”³ De grundlæggende begreber i infologien er data og information. Disse to ord kan defineres på mange måder, men jeg vil undlade en diskussion og sammenligning af de forskellige definitioner og holde mig til Langefors’ forklaringer.

1.1. Data, information og den infologiske ligning

Data er ikke information og indeholder ikke information, men data repræsenterer information ved hjælp af tegn. Det er begrebet repræsentation, der er det centrale. Selv om Langefors beskæftiger sig med edb-systemer, pointerer han, at begrebet data ikke kun skal forstås i relation til en computer. Data kan også være en skrevet sætning, hvor tegnene er en repræsentation af information.⁴ Derimod beskæftiger Langefors sig ikke med information, der formidles på anden vis end med skrift, og det vil jeg heller ikke gøre. Arkivalier er ganske vist ikke nødvendigvis skriftlige (elektroniske eller papirbårne), men det vil føre for vidt her at komme ind på fx billeder eller audiovisuelle arkivalier.

Langefors har ikke en egentlig definition af information, men forstår det tilsyneladende som stort set synonymt med viden, dog med den tilføjelse at information er viden, der kan kommunikeres og lagres (på papir eller elektronisk).⁵ Han anvender også udtrykket eksplicit viden som synonymt med information og modstiller det implicit viden. Implicit viden findes i en persons hoved og er ikke information. Hvis denne viden kan struktureres og repræsenteres ved hjælp af data, bliver den eksplicit viden og information. Det vil sige, at der i begrebet information medtænkes et kommunikationsaspekt.

Der er en grundlæggende forskel på data og information, samtidig med at de er indbyrdes forbundne, og Langefors taler om ”the infological dichotomy”⁶. Information findes i en persons hoved, mens data er en fysisk repræsentation af information. Spørgsmålet

³ Langefors (1981), s. 66

⁴ Fx Langefors (1995), s. 108

⁵ Langefors (1995), s. 107

⁶ Fx Langefors (1977), s. 7

er så, hvordan transformationen fra data til information foregår, og det kan beskrives ved den såkaldte infologiske ligning:

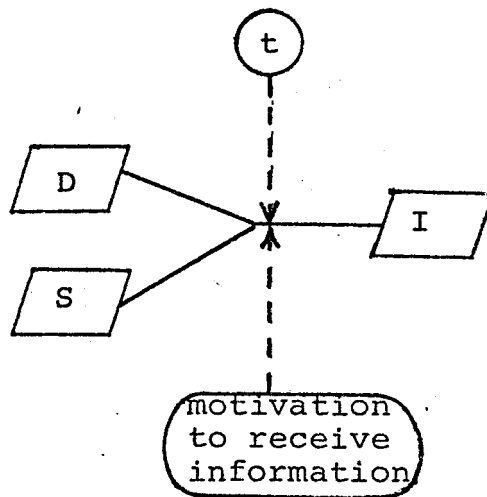
$$I = i (D, S, t)^7$$

I står for information, og D for data, begge som beskrevet ovenfor. Det lille i står for interpretation, det vil sige den proces, hvor data bliver fortolket til information (fortolkningsprocessen). I denne proces indgår også S og t.

Hvad S er en forkortelse for, er lidt uklart, men jeg antager, at det står for "semantic background". Det er lidt forskelligt, hvilken benævnelse Langefors anvender, men jeg vil i det følgende kalde det for forudviden (foreknowledge).⁸ Med forudviden menes den viden, som en fortolker besidder, og som benyttes i fortolkningsprocessen.

I det følgende kaldes den person, der foretager fortolkningsprocessen for fortolkeren, mens Langefors bl.a. bruger det upersonlige udtryk "den modtagende struktur" (the receiving structure).⁹ Dette upersonlige udtryk åbner for overvejelser om, hvorvidt modtageren eller fortolkeren kan være en computer, men derved bevæger man sig ind på problemkomplekset vedrørende kunstig intelligens, og det vil jeg afholde mig fra i denne sammenhæng.

Det lille t står for tid, og det kan tilsyneladende enten være det tidsrum, indenfor hvilken fortolkningsprocessen forløber, eller det tidspunkt hvor fortolkningen foregår. Jeg vil i det følgende definere t på den sidstnævnte måde, jf. figur 1, som skal illustrere, at data kan eksistere i meget lang tid som data. Det er først, når data på et givet tidspunkt i en fortolkningsproces bringes sammen med fortolkerens forudviden, at der skabes information.



Figur 1.¹⁰

⁷ Læses på følgende måde: Information opstår ved en fortolkningsproces, hvori indgår data, forudviden og tid.

⁸ Langefors (1977), s. 12

⁹ Langefors (1977), s. 6

¹⁰ Langefors (1977), s. 6

Jeg skal forsøge at anskueliggøre forskellen mellem data og information med et simpelt eksempel, nemlig ordet 'internet'. Disse otte små tegn sat sammen i denne rækkefølge repræsenterer en information om et fænomen i virkelighedens verden. Med dette fænomen i virkelighedens verden dukker der imidlertid et nyt niveau op, som i det følgende vil blive kaldt informationsobjektet. Det er et objekt, fysisk eller tænkt, en situation, et eller andet der eksisterer i virkelighedens verden.¹¹

Når tænkt objekt nævnes, er der dog noget, der ikke hænger sammen, for kan et tænkt objekt være et informationsobjekt, når det ikke eksisterer fysisk, men kun i vore hoveder. Langefors anvender også i stedet for informationsobjekt begrebet elementær situation. Dette begreb kan synes bredere, da det ikke giver et billede af noget fysisk eksisterende (som objekt gør), og da det også giver et billede af, at der sker noget, og at der er nogle attributer tilknyttet.¹² Jeg vil dog i det følgende benytte begrebet informationsobjekt, men blot gøre opmærksom på, at det skal forstås bredt, som fysisk objekt, tænkt objekt (begreb), hændelse eller situation, eksisterende i verden, og med nogle attributer tilknyttet, fx egenskaber eller handlinger.

Tilbage til det lille eksempel med 'internet'. Disse otte bogstaver er data, der af læseren fortolkes til information om et informationsobjekt. Data er tegn på et stykke papir, information er i fortolkerens hoved, og informationsobjektet findes i virkelighedens verden.

Forudsætningen for at fortolkningen lykkes, er flere typer af forudviden, først og fremmest at fortolkeren kan læse, og dernæst at han kender fænomenet. Ordet 'internet' kunne også læses for 20 år siden, men det ville ikke kunne fortolkes til information, da ingen havde forudviden om et fænomen, der ikke eksisterede på dette tidspunkt.

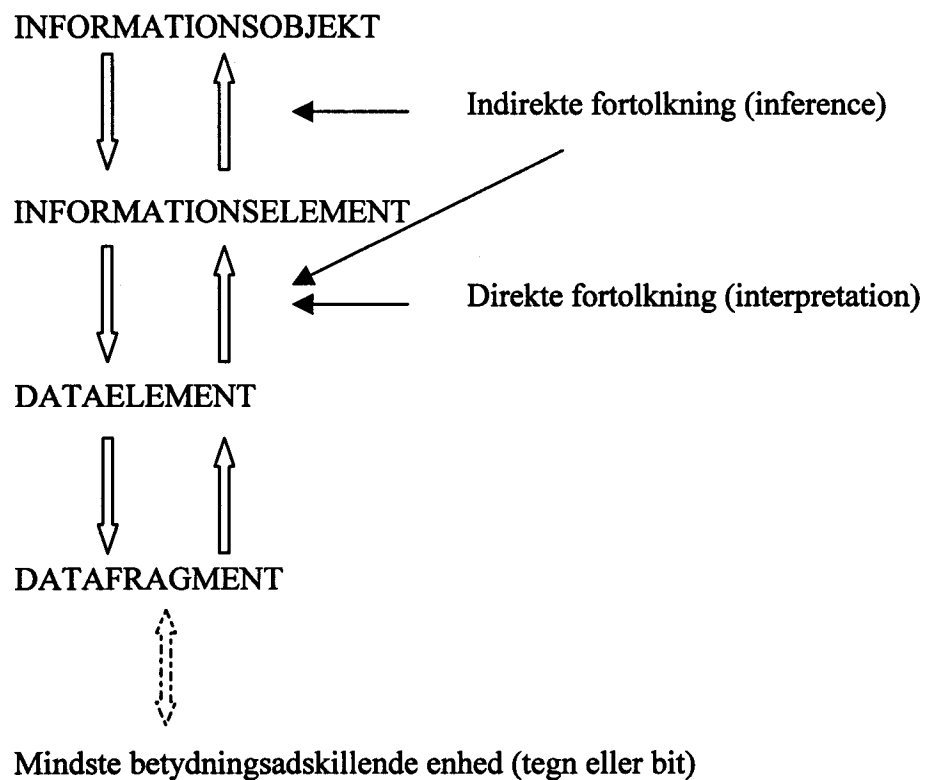
Et eksempel på at kendskab til skriftsproget og det anvendte alfabet er en forudsætning for at afkode data og dernæst fortolke data til information, kan være en russisk fortolker, der kun kender det kyrilliske alfabet. Han kender fænomenet internet, og han kender betegnelsen for fænomenet, der er den samme på russisk. Men da han ikke kan tyde de latinske bogstaver, kan han alligevel ikke fortolke data.

1.2. Den infologiske model

Den infologiske model er baseret på Langefors og en række af hans begreber, men modellen som vist i figur 2 er mit forsøg på at illustrere nogle sammenhæng. Jeg har desuden tilføjet modellen begreberne datafragment og mindste betydningsadskillende enhed, som ikke indgår i Langefors' begrebsapparat.

¹¹ Langefors (1995), s. 110

¹² Langefors (1976), s. 57

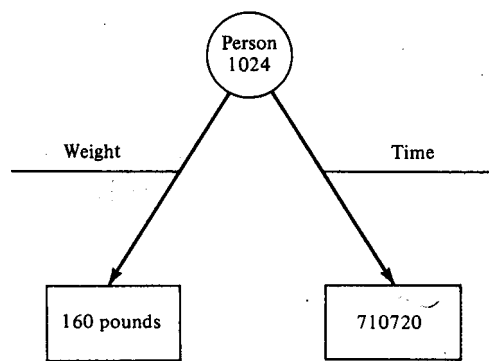


Figur 2

Informationsobjektet er som beskrevet ovenfor noget, der eksisterer i verden: et objekt, et begreb, en hændelse eller en situation. Informationsobjektet kan både være fysisk og abstrakt. En journal indeholder data, der er en repræsentation af information om eksistensen af en sag (informationsobjektet). Selve sagen indeholder data, der er en repræsentation af information om den hændelse, at en virksomhed har søgt tilskud til udvikling af nye energikilder (informationsobjektet). Denne hændelse (søgning af tilskud) er passeret, så informationsobjektet eksisterer ikke mere. Det vil sige, at et informationsobjekt ikke nødvendigvis eksisterer i verden lige nu. Men der er registreret data, der indeholder information om dets eksistens på et tidspunkt før nu'et, og derfor ved vi, at dette informationsobjekt har været.

Ovenfor har jeg forsøgt at eksemplificere forskellen på data og information med 'internet', men faktisk repræsenterer dette ord isoleret en ufuldstændig information, for hvad er der med det internet? Vil jeg fortælle, at det eksisterer (hvornår), at det er interessant, at der kan findes information ved hjælp af det eller noget helt fjerde?

Langefors anvender i nogle sammenhænge i stedet for begrebet information udtrykket informationselement med den hensigt at pointere, at fuldstændig information om et informationsobjekt kan beskrives som bestående af seks dele. Informationselementet findes som sagt i fortolkerens hoved, men kan repræsenteres ved et dataelement, der tilsvarende består af seks dele. Langefors anvender nedenstående figur til illustration af de seks dele, som et dataelement består af:



Figur 3.¹³

- Objektkategori og identifikation af objektet inden for kategorien (person og 1024).
- En attribut og værdien af attributen (inkl. evt. angivelse af skala) (weight og 160 pounds). Attribut skal forstås meget bredt som et eller andet, der karakteriserer objektet, og det kan også være angivelse af relation til et andet objekt. Ofte vil der i praksis være knyttet flere attributer til et objekt, men det vil altid kunne nedbrydes til et såkaldt elementært dataelement med kun én attribut.
- Angivelse af tid og en specifikation af tidspunktet eller tidsintervallet (time og 710720).

Men et dataelement behøver ikke nødvendigvis at indeholde alle seks dele for at kunne give en fortolker et fuldstændigt informationselement. Dele af informationselementet kan være underforstået, fordi den, der skaber dataelementet for at repræsentere informationselementet, ved eller antager, at fortolkeren har en forudviden.¹⁴

Hvis vi antager, at eksemplet i figur 3 illustrerer et led i en større videnskabelig undersøgelse af en række personer, kunne angivelsen af objektkategorien (person) udelades og ligeledes skalaangivelsen (pound), for det antages at være kendt af fortolkerne. Hvis alle forsøgspersoner blev vejet samme dag, kunne tidsangivelsen ligeledes udelades fra den enkelte registrering og angives ét sted (fx som overskrift på en liste).

Mindre væsentligt i forhold til arkivalier (men ikke i forhold til systemudviklingen og udviklingen af brugergrænseflader, der er i fokus for Langefors) er, at et informationselement kan repræsenteres af forskellige dataelementer. Fx kan informationen om en persons vægt repræsenteres med figuren ovenfor, men den kunne også repræsenteres med en sætning.

Som en illustration af at et dataelement består af seks dele, har jeg på den infologiske model i figur 2 angivet datafragment, der ikke er et begreb, som Langefors benytter. Datafragment er fx '1024' eller '710720' i eksemplet i figur 3 eller det tidligere nævnte ord 'internet'. Et datafragment repræsenterer ikke information, før det bliver koblet sammen med andre datafragmenter til et dataelement. I en sætning kan datafragmentet

¹³ Langefors (1976), s. 53

¹⁴ Langefors (1977), s. 8 - 9

være det enkelte ord, som normalt ikke giver en fuldstændig information, og da slet ikke hvis ordet kan have flere betydninger afhængig af sammenhæng.

Nederst på den infologiske figur har jeg angivet 'mindste betydningsadskillende enhed', dog med lille skrift og med en stiplede pil for at angive, at jeg kun i begrænset grad vil beskæftige mig med dette niveau. Men det enkelte tegn i skrevet tekst er betydningsadskillende og dermed afgørende for datafragmentet og resten af modellen. Det betyder noget for informationselementet, om der står 'Han tager en lur' eller 'Han tager en tur'.

En af de helt grundlæggende forskelle på lagring af data på papir og på elektroniske medier er, at når mediet er papir, er tegnet, fx t eller l, den mindste betydningsadskillende enhed. Når det gælder elektronisk lagrede data, skal man helt ned på bit-niveau for at finde forskellen, der gør en forskel – og det er ikke synligt for det menneskelige øje.

1.3. Fortolkning og forudviden

På den infologiske model har jeg lavet pile begge veje mellem de enkelte niveauer, men her vil jeg først og fremmest fokusere på den opadgående pil mellem dataelement og informationselement, som også er den transformation, Langefors har fokus på.

For at omforme et dataelement til et informationselement skal der ifølge den infologiske ligning ske en fortolkningsproces, hvori indgår fortolkerens forudviden. Hvis vi igen tager det lille eksempel med vejningen i figur 3, skal fortolkeren for det første kende tegnene (alfabetet) og det anvendte sprog. Han skal også kende vægtskalaen (jeg ved, at jeg skal have regnet pound om til kilogram for at få information), og han skal vide, hvordan datoen skal læses (er det år, der kommer først?). Så langt kan jeg som fortolker godt følge med, men jeg får egentlig ikke nogen brugbar information, for hvem er denne person, hvorfor skal vedkommende vejes, og hvad skal informationen bruges til? Jeg kan forestille mig meget, men jeg aner det ikke, fordi jeg ikke kender konteksten, og derfor er det ikke brugbar information for mig.

Langefors definerer forudviden som al den viden, fortolkeren har samlet sammen i sit hoved gennem et helt liv, og det får ham til at diskutere det interessante filosofiske problem, om vi overhovedet kan få den samme information ud af data, og om vi overhovedet kan kommunikere, eftersom hvert enkelt menneskes forudviden er forskellig. Han konstaterer, at han har set mange eksempler på, at kommunikation er slået fejl, men at han også har set mange eksempler på det modsatte, og at kommunikation, hvor personer får samme information ud af data, især synes at fungere inden for mindre grupper, der har et eller andet fællesskab, fx et fagligt fællesskab.¹⁵

Derfor konkluderer Langefors, at en gruppe mennesker kan have delvis fælles forudviden, betinget af at de hører til det samme kommunikationsfællesskab, og at fortolkningsprocessen må deles op i to.¹⁶

Den direkte fortolkning (interpretation) medfører en faktuel information om informationsobjektet, og til denne fortolkning anvendes en del af S, kaldet observations-

¹⁵ Langefors (1995), s. 144 - 145

¹⁶ Langefors (1995), s. 152 - 153 og Langefors (1981), s. 73 - 74

forudviden, men jeg vil foretrække at kalde den fælles forudviden. Personer, der er deltagere i et kommunikationsfællesskab, antages at få samme faktuelle information ud af et dataelement. Fx vil en gruppe, der i videnskabeligt øjemed foretager jævnlige vejninger af en gruppe mennesker, få samme information ud af vægt-eksemplet ovenfor, og de vil også have den nødvendige viden om konteksten, som jeg mangler. Langefors nævner det ikke, men implicit i kommunikationsfællesskab ligger ikke blot en forudsætning for at forstå ordene (dataelementerne) ens, men også en viden om konteksten.

Den indirekte fortolkning (inference) kommer efter den direkte fortolkning, og den er individuel, da den er baseret på fortolkerens samlede S, kaldet slutnings-forudviden, men jeg vil foretrække at kalde den privat forudviden. I gruppen, der vejer forsøgspersoner, er en sygeplejerske, der foretager vejningerne og registreringer. Da hun ser registreringen om personen 1024, tænker hun: ”Pyha, han er blevet vejjet, så jeg slipper for ham, for han er da en irriterende personage.” I gruppen er også en læge, der tænker: ”Jeg må hellere finde ud af alder, køn og højde for at kunne vurdere, om personen er normalvægtig – er det ikke den lille dame på 150 cm”. Vægtvogteren, der også har fået sneget sig ind i gruppen, tænker: ”Pokers også, der er vist ikke nogen kunde her.”

Hvis begreberne forudviden og fortolkning forsøges overført til arkivverdenen og brug af kilder, er den fælles forudviden den viden, som er fælles for en gruppe af forskere på grund af deres uddannelsesmæssige baggrund og tilhørsforhold til et fagligt fællesskab. Den direkte fortolkning af kilderne fører hos dem til samme information, f.eks. når det gælder faktuelle begivenhederne. Men når forskerne forsøger at forklare årsager til og sammenhænge mellem begivenheder, sker der en indirekte fortolkning baseret på den private forudviden, der fører til individuelle valg af spørgsmål og metoder. Groft sagt kan man sige, at et arkiv er med til at sikre den nødvendige forudviden for den direkte fortolkning, mens den indirekte fortolkning ikke vedrører arkivet.

Opdelingen i fælles og privat forudviden kan synes kunstig, for hvor er grænsen mellem de to, både teoretisk og i den enkelte fortolkningsproces. Den kan man ikke finde, for hvem skulle gøre det. Fortolkeren selv kan ikke, for han kan ikke vide, præcis hvilken forudviden han har tilfælles med de øvrige i kommunikationsfællesskabet (der vil nok også være variationer fra person til person). En udenforstående observatør kan slet ikke, for han har ikke adgang til fortolkerens forudviden. Så opdelingen af forudviden skal ikke ses som en praktisk mulighed for analyse, men som en erkendelse af, at ethvert menneske har en privat og individuel forudviden, men en del af denne viden kan være fælles med en gruppe mennesker, hvor den fælles viden udvikles gennem tilhørsforholdet til et kommunikationsfællesskab.

2. ARKIVALIER SET FRA EN INFOLOGISK OG EN ARKIVISTISK SYNSVINKEL

Formålet med det foregående afsnits gennemgang af den infologiske ligning og den infologiske model er som før nævnt at skabe en ramme, inden for hvilken nedenstående kan diskuteres:

- Nogle specifikke kendetegn ved arkivalier i forhold til anden information
- Forskelle på arkivalier på papir og e-arkivalier
- Forskelle på IT-systemer i brug hos en arkivskaber og arkiveringsversioner af IT-systemerne.

2.1. Begrebet arkivalier

Jeg vil indlede med nogle overvejelser over, hvad der traditionelt forstås ved arkivalier. Det har imidlertid vist sig svært at finde definitioner på begrebet, måske fordi det er indlysende, at det – som det defineres i Fremmedordbogen – er dokumenter, der opbevares i et arkiv. Jeg har dog fundet to definitioner. For det første er der en definition i Arkivhåndbogen:

”Et arkivalie er et medium (papir, CD-rom, film mv.), der indeholder informationer, som er blevet opsamlet i forbindelse med en myndigheds virksomhed.”¹⁷

For det andet er der ICA's definition. Her skal det lige indskydes, at det engelske ord for arkivalie - record - er et godt eksempel på et udtryk, der kan vække megen forvirring, da betydningen afhænger af, hvilket kommunikationsfællesskab en fortolker tilhører. Det kan især gå galt, når personer med arkivbaggrund taler om 'electronic record', for en person med informationsteorisk eller datalogisk baggrund vil forstå det som en post i en database. Definitionen er som følger:

”A record is recorded information produced or received in the initiation, conduct or completion of an institutional or individual activity and that comprises content, context and structure sufficient to provide evidence of the activity.”¹⁸

I den danske definition er arkivalier indskrænket til at være skabt af myndigheder, hvilket må antages at skyldes, at Arkivhåndbogen er rettet mod myndigheder. Det menes næppe, at arkivalier ikke kan være skabt af virksomheder, organisationer eller private. Begge definitioner er enige om, at det drejer sig om informationer, der er opsamlet eller registreret i forbindelse med aktivitet. Den danske definition har fokus på mediet, mens ICA's definition fokuserer på, at der er behov for både indhold, kontekst og struktur for at give tilstrækkeligt vidnesbyrd om aktiviteten.¹⁹

¹⁷ Arkivhåndbogen

¹⁸ Guide for Managing Electronic Records, afsnit 2.1.

¹⁹ Evidence skal oversættes som vidnesbyrd.

Det sidste årti har der internationalt været en del diskussionsindlæg om 'elektroniske records', altså e-arkivalier, og hvordan de kan defineres. I Guiden fra 1997 anvendes ovenstående ICA-definition også på elektroniske arkivalier, hvorefter det beskrives, hvorledes disse adskiller sig fra papirbundne arkivalier:

- Selve lagringen sker ved hjælp af symboler, der ikke umiddelbart kan læses af mennesker.
- Indhold og medie kan adskilles.
- Den fysiske struktur er ikke kendt af brugeren og uvæsentlig i forhold til den logiske struktur: "This logical structure of an electronic record will, generally speaking, often be the structure which the producer created on his screen. In order to be considered complete and authentic, the record must preserve this structure in some way, and the computer system must reconstruct it when transforming the record back to a human-readable format."²⁰

Guiden gør dermed opmærksom på, at den generelle ICA-definition på den ene side er bred nok til at omfatte elektroniske arkivalier, men at der på den anden side er nogle helt afgørende forskelle. Hvis jeg skal sammenfatte mit indtryk af diskussionsindlæggene det sidste årti, synes disse forskelle næsten at have skræmt arkiv-samfundet, hvor bekymringerne for mig at se har kredset om følgende punkter:²¹

- De store teknologiske problemer, hvor tendensen oprindelig var, at man gerne ville bevare arkivalierne, som de var, ved at anvende den museale strategi eller emuleringsstrategien. Det sidste par år er tendensen i praksis tilsyneladende gået imod bevarelse ved konvertering til ikke-proprietære formater.
- Frygten for manglende arkivalier, ikke på grund af teknologisk forældelse, men fordi dokumenter ikke længere vil være pænt samlet i et arkiv, men være lagret hist og pist i systemer, der ikke umiddelbart er erkendbar med det menneskelige øje.
- Bekymring for arkivernes (og arkivarernes) fremtid, fordi det forekommer vanskeligt at afgrænse digitale arkivalier fra al anden digitalt lagret information. Det rejser følgelig spørgsmålet om, hvorvidt det ikke lige så godt kan være biblioteker (og bibliotekarer) eller document managers, der i fremtiden har ansvaret for digitalt lagrede arkivalier.
- Problemerne med at få hold på fænomenet databaser. I Guiden fra 1997 anføres det som nævnt, at den generelle definition på arkivalier også kan anvendes på elektroniske arkivalier, men det nævnes også, at det er nemmest, når det gælder dokumenter, men mere problematisk ved databaser.²² Det diskuteres fx, om en database er ét arkivalie eller mange, og hvis det er mange, hvordan det enkelte arkivalie så skal identificeres. I en del tilfælde forsøges problemet afvist ved simpelthen at hævde, at databaser ikke er arkivalier, blot samlinger af information uden interesse for arkiver.

²⁰ Guide for Managing Electronic Records, afsnit 2.2.

²¹ Mit indtryk er baseret på læsning af den arkivistiske litteratur, bl.a. The Concept of Record samt Erlandsson.

²² Guide for Managing Electronic Records, part II, afsnit 1.2.

Der er ret bred enighed om, at risikoen for ufuldstændig arkivdannelse og overflødiggørelse af arkivarer skal imødegås ved, at arkivarer holder op med at betragte arkivalier som noget, der først angår dem, når arkivalierne er gået ud af brug. Det er en nødvendighed, at arkivarerne er aktivt involverede, allerede når et arkivalie skabes eller allerede i forbindelse med systemudviklingen, så det sikres, at ”records are captured”. Som supplement til de traditionelle begreber custodial og non-custodial indføres begreberne post-custodial og record continuum.

Med begrebet post-custodial menes, at arkivarer skal holde op med at være kustoder forskanset bag tykke mure og i stedet blande sig i arkivdannelsen lige fra starten, uden at det har noget at gøre med, om arkiverne skal forblive i skabernes varetægt eller afleveres til arkiv. Formålet med begrebet record continuum er ligeledes at understrege, at et arkivalie ikke først skal varetages af arkivarer ved aflevering til arkiv, men skal varetages allerede fra før dets opståen, samt at der ikke er noget skel mellem aktuel administrativ brug og historisk brug. Det sidste er langt fra alle dog enige i.

De nærmere beskrivelser af, hvordan involveringen skal praktiseres, varierer afhængig af nationale traditioner og juridiske muligheder. Der er dog også kritiske røster imod indblanding ud fra den betragtning, at hvis arkivarer påvirker arkivdannelsen af hensyn til den historiske brug, vil den oprindelige administrative brug ikke blive korrekt afspejlet i arkivdannelsen.

Det retter opmærksomheden imod spørgsmålet om, hvornår et arkivalie egentlig opstår. Opstår det i samme øjeblik tegn fæstes til papiret eller lagres elektronisk? Opstår det allerede inden, når det planlægges, at der i forbindelse med en given aktivitet skal fæstes tegn til papir eller lagres elektronisk? Opstår arkivaliet først, når data går ud af administrativ brug – hvis det overhovedet kan fastsættes præcist, jævnfør begrebet ’semi-active records’? Opstår arkivaliet først, når det afleveres til arkiv?

Ud fra en praktisk og juridisk synsvinkel kan det være hensigtsmæssigt at argumentere for, at arkivalier opstår straks ved deres skabelse, eller allerede når de planlægges, så det sikres, at arkiverne kan påvirke arkivdannelse ud fra hensynet til brugen på længere og meget langt sigt.

Men ud fra min teoretiske indgangsvinkel vil jeg mene, at arkivalier først opstår, når papirerne med tegn eller de digitalt lagrede data løsrives fra deres oprindelse i en organisation og et kommunikationsfællesskab og afleveres til arkiv. I Schellenbergske arkivtermer vil det sige, at arkivaliet først opstår, når data får sin sekundære værdi eller genanvendelsesværdi, ikke når data er i brug på grund af sin primære eller oprindelige værdi.²³

2.2. Data, information og arkivalier

En myndighed eller en organisation har i udøvelse af sin virksomhed behov for at fastholde viden om informationsobjekter, det være sig fysiske objekter, situationer, hændel-

²³ Jf. Schellenberg, s. 6. Schellenberg bruger begreberne primary og secondary values, der i ”Kassation og bevaring af arkivalier” oversættes som henholdsvis oprindelig værdi og genanvendelsesværdi (s. 17).

ser eller abstraktioner. Behovet skyldes både nødvendigheden af dokumentation udadtil og den omstændighed, at en organisation består af mere end en person, og viden derfor ikke kan forblive gemt som implicit viden i den enkelte medarbejders hoved, men skal repræsenteres som dataelementer, så de øvrige medarbejdere har adgang til den. Hvis man ser på den infologiske model (figur 2), illustreres omformningen fra implicit viden til data med de nedadgående pile mellem informationsobjekt, informationselement og dataelement. Det kan også beskrives med, hvad jeg vil kalde den omvendte infologiske ligning:

$$D = i(I, \text{Sobs}'A, t).^{24}$$

Data skabes ved den omvendte fortolkningsproces, hvori udover tiden indgår informationen og den fælles forudviden (Sobs'A), der er knyttet til det kommunikationsfællesskab, som findes i organisationen. Der er selvfølgelig en risiko for, at dataskaberer bruger noget af sin private forudviden, men det vil normalt hurtigt blive rettet, fordi fortolkeren kender dataskaberer og vil gøre opmærksom på, at der er et problem.

Data bliver som sagt skabt, fordi andre i organisationen til interne formål eller til dokumentation udadtil har brug for den information, de repræsenterer, og der vil derfor finde en eller mange fortolkningsprocesser sted:

$$I = i(D, \text{Sobs}'A, t),$$

hvor Sobs'A som sagt er den fælles forudviden, der er identisk for skaber og fortolkerne af data.

På et tidspunkt er den information, som data repræsenterer, ikke længere nødvendig for organisationen hverken til interne eller eksterne formål. Hvis organisationen er omfattet af afleveringspligt til arkiv eller selv vælger at aflevere, og hvis de informationerne om informationsobjekter, som data repræsenterer, vurderes at skulle bevares af historiske eller dokumentariske hensyn, overføres data til arkiv.

Her vil data igen før eller siden indgå i en fortolkningsproces:

$$I = i(D, \text{Sobs}'B, t).$$

D er principielt uforandret. Det er dog ikke helt givet, for det kræver, at arkivet sikrer sig, at hverken bits eller tegn på papiret ændrer sig, hvilket kan kræve henholdsvis mediekonvertering og konservering. Men t er en helt anden, og det er S også på grund af tidsforskydningen og på grund af, at fortolkeren ikke er en del af det oprindelige kommunikationsfællesskab – og heller ikke kan blive det, da det ikke længere findes. Arkivfortolkeren må derfor forsøge at tilnærme sin forudviden (Sobs'B) til den fælles forudviden i den oprindelige kommunikationsfællesskab (Sobs'A) ved at analysere den kontekst, som data er skabt i.

²⁴ Læses på følgende måde: Data opstår ved en fortolkningsproces, hvori indgår information, forudviden og tid.

Det kræver kilder til konteksten, og de vil ofte findes, fordi et arkivalie normalt er skabt i en velstruktureret organisatorisk sammenhæng. En del af det oprindelige kommunikationsfællesskab vil ganske vist være bundet til det uformelle, det sociale, sprog-spillene, den forudviden, der kun har eksisteret ustruktureret og uformuleret i medlemmernes hoveder, og denne del af forudviden er borte. Men andet vil findes nedskrevet eller ligge implicit i arkivaliet.

For det første vil det enkelte dataelement ikke optræde løsrevet, men indgå i sammenhæng med andre dataelementer. Et enkelt dokument vil ikke ligge løsrevet, men på en eller anden måde være kædet sammen med andre dokumenter. En indførsel i en kirkebog vil stå sammen med andre tilsvarende indførsler og være sat ind i kirkebogens struktur. I denne struktur og sammenhæng med andre dataelementer ligger implicit en del af forudviden.

For det andet vil samlingen af sammenhængende dataelementer – også kaldet arkivserie eller heuristisk enhed – være beskrevet af arkiverne. Det lader til gennem mange år at have været anset for en væsentlig opgave for arkiverne at sikre oplysninger om proveniens og den organisatoriske kontekst samt sammenhæng med andre arkivserier, fordi det ikke ligger implicit i arkivaliet – konteksten var på skabelsestidspunktet en del af den fælles forudviden.

For det tredje vil en del af forudviden findes overleveret i andre kilder – hvis de er bevaret, og hvis man kan finde dem.

Sammenfattende kan siges, at arkivalier adskiller sig fra andre informationsrepræsenterende data ved at være løsrevet fra det organisatorisk begrænsede kommunikationsfællesskab, som data er skabt i og beregnet til at blive anvendt inden for. Dette gælder, uanset om data er repræsenteret ved tegn på papir eller i digital form. Arkivernes opgave er at sikre så megen forudviden fra kommunikationsfællesskabet som muligt, det vil sige struktur og kontekst samt – mere diffust – andre kilder, der indeholder forudviden af relevans for en fortolkning af data.

I forrige afsnit er nævnt de internationale overvejelser om forholdet mellem arkiver og biblioteker. Set fra en infologisk synsvinkel er en central forskel, at det publicerede materiale i et bibliotek retter sig mod et mere alment kommunikationsfællesskab end arkivalier.²⁵ Den information, der er repræsenteret i arkivalier, er som nævnt rettet mod en snæver veldefineret organisatorisk kontekst, mens information, der er repræsenteret i publikationer, gerne skal spredes så meget som muligt. Ved arkivalier giver det mening at anvende de Schellenbergske termer oprindelig værdi og genanvendelsesværdi, men det giver ikke mening ved publikationer.

²⁵ Hvor alment afhænger naturligvis af arten af publikationen. Der er fx forskel på forfatterens antagelse om kommunikationsfællesskab ved en artikel i et dagblad og i et fagtidsskrift.

2.3. Proveniensbegrebet og elektroniske arkivalier

Proveniensbegrebet står som et helt centralt begreb i arkivstikken, og udover begrebet arkivalie eller record er det det begreb, som mange peger på, når de skal argumentere for arkivernes berettigelse i en fremtidig elektronisk verden.²⁶

Ordet proveniens betyder egentlig oprindelse eller være kommende fra (af latin provenire: komme fra, stamme fra). Men spørgsmålet er selvfølgelig, hvordan proveniensprincippet defineres i et arkivalsk kommunikationsfællesskab. Det er der skrevet en del om de sidste godt 150 år, og tilsyneladende er der en vis enighed om en grundlæggende definition af begrebet. Men når der går mere i detaljer både teoretisk og i praksis, er der forskellige fortolkninger afhængig af tid, geografi og de enkelte personer. Jeg vil ikke her gå ind i en beskrivelse af udviklingen og de geografiske forskelle, men blot give en grov sammenfatning med det formål at holde princippet op imod e-arkivalier.

Proveniensprincippet blev opfundet i praksis i løbet af 1800-tallet i en række europæiske lande mere eller mindre uafhængigt af hinanden. Inden da havde man ofte forsøgt at ordne arkivalierne efter emne, det vil sige efter et bibliotekslignende princip helt ned på det enkelte dokumentets niveau. Det viste sig imidlertid efterhånden at være meget vanskeligt og i praksis uoverkommeligt. Men et eller andet ordningsprincip var der behov for både for at kunne placere arkivalierne på en struktureret måde på hylderne og for at kunne lave oversigter over bestanden. Så da ordningen efter emne blev for kompliceret, var der behov for et andet princip.

Proveniensprincippet blev i 1898 beskrevet i *The Dutch Manual* af Muller, Feith og Fruin, og det er normalt dette værk, der refereres til som det teoretiske grundlag. I de enkelte lande har princippet som sagt over tid både teoretisk og i praksis udviklet sig forskelligt, men grundlæggende består princippet af to dele: at en arkivskabers fond ikke må sammenblandes med en anden arkivskabers fond (*respect des fonds*), og at arkivaliernes indre orden ikke må ændres (*respect de l'ordre interieure*). Der er dog en vis uenighed om, hvorvidt respekten for den indre orden er en del af proveniensprincippet, eller om det er et princip ved siden af proveniensprincippet, der så kun drejer sig om respekten for fonden.

Det første princip om respekten for fonden gjorde det meget nemmere at placere arkivalierne på hylderne og også at beskrive dem i registraturer, fordi de blev beskrevet som et hierarki, hvor det øverste niveau var den organisation eller person, der havde skabt arkivaliet eller afleveret det. I spørgsmålet om skabt eller afleveret ligger for øvrigt kimen til en af mange diskussioner om proveniensprincippet, for hvis respekten for fonden ikke er blevet respekteret inden aflevering til arkiv, er spørgsmålet, om arkivet skal fjernes de dele, der er skabt af en anden end den, der afleverer, og indordne dem i det oprindelige arkiv. Ligeledes har arkiver haft problemet med de arkivalier, der allerede var blevet ordnet efter et emneprincip, for kunne og skulle de ordnes tilbage efter proveniensprincippet. Ud over de praktiske argumenter for ordning efter proveniensprincippet

²⁶ Den kortfattede gennemgang af proveniensbegrebet er baseret på det overordnede indtryk, jeg har fået ved læsning af den arkivistiske litteratur, der fremgår af litteraturlisten, først og fremmest *The Principle of Provenance*.

har der også været andre argumenter, fx at det for historieforskere viste sig frugtbart at se arkivalierne i den sammenhæng, de er skabt i. Dette vil jeg dog ikke komme nærmere ind på, da det også hænger sammen med en længere redegørelse for historiefagets udvikling.

Der har tilsyneladende ikke været sat mange spørgsmålstejn ved proveniensprincippet som sådan, kun dets nærmere fortolkning. Men ordningen efter emne havde også sine fordele, og der var og er stadig behov for en sådan ordning som grundlag for fremfindning af arkivalier, hvilket kan ses på flere måder. For det første er der de anekdoteliggende historier om ikke arkivvante brugere, der kommer ind og beder om at få alt materiale om fx et lokalområde – vel egentlig en meget naturlig forventning i vore tider, hvor mulighederne for informationssøgning er legio. For det andet forekommer det, at højfrekvente arkivalier udtages fra en fond og placeres sammen med tilsvarende arkivalier fra andre fonde. For det tredje kan det ses af de forsøg, der gøres på at lave emnebaserede indgange til arkivdatabaser.

Det andet princip om respekten for den indre orden synes også praktisk begrundet, idet det vil være meget u hensigtsmæssigt at omordne en sammenhængende samling af arkivalier, hvis de har en orden, som gør det muligt at finde rundt i dem. Arkivskaberen kan have ordnet arkivalierne i en fysisk rækkefølge, der kan beskrives, og som gør det nemt at finde de enkelte dokumenter, eller arkivskaberen har selv haft et søgemiddel i form af fx et register eller en journal. Når man går mere i detaljer, er der også her uenighed, idet det ene yderpunkt er, at den indre orden er hellig og ikke må krænkes. Det andet yderpunkt er, at hvis den indre orden er fraværende eller uforståelig for andre end skaberen, er det både acceptabelt og fornuftigt at skabe en ny indre orden.

Som nævnt er der også i dag stor enighed om proveniensprincippet som det unikke grundlag for arkiverne. Den bredest accepterede, nugældende definition må antages at være ICA's:

“Provenance (is) the relationships between records and the organizations or individuals that created, accumulated and/or maintained and used them in the conduct of personal or corporate activity.”²⁷

Det er en meget bred definition, hvor fokus er på relationerne mellem arkivalier og organisationer (eller personer), og der tages ikke stilling til karakteren af disse relationer, idet organisationerne kan have skabt eller modtaget arkivalierne, men de kan også blot have brugt dem eller have haft dem i varetægt. Med hensyn til fokus på relationer har jeg forøvrig et indtryk af, at der her er sket et afgørende skift væk fra en hierarkisk betragtningsmåde – men jeg er ikke gået tilstrækkelig i detaljer til at kunne dokumentere det.

Spørgsmålet er imidlertid, hvorvidt proveniensbegrebet teoretisk og i praksis er brugbart i forhold til e-arkivalier.

Den del af proveniensprincippet, der drejer sig om den indre orden, er vanskelig at anvende i forhold til e-arkivalier. Som det vil fremgå mere detaljeret af det følgende afsnit

²⁷ ISAAR(CPF), s. 10

2.4 , er indhold og struktur adskillelige i IT-systemer, og de vil uundgåeligt blive adskilt i det øjeblik, der laves en systemuafhængig arkiveringsversion. Det er ikke, fordi der ikke er indre orden i såvel et IT-system som en arkiveringsversion. Faktisk er der mange indre ordener. Der er en indre orden i den enkelte fil eller tabel, imellem tabellerne, hvis det er en relationel database, og en indre logisk orden i den kombination af data, der præsenteres for brugeren. Det er væsentligt at være opmærksom på, i hvilket omfang der er behov for bevaring af alle disse indre ordener, men jeg finder, at det bliver søgt og ufrugtbart at forsøge at tilpasse princippet yderligere til e-arkivalier.

Proveniensprincippet i betydningen respekt for fonden eller oprindelsen er derimod væsentligt. Som det fremgår af det foregående afsnit, er det ud fra en infologisk analyse centralt for en fortolkning af data at kende kommunikationsfællesskabet, som er snævert forbundet med den organisatoriske kontekst og oprindelsen. Jeg har ovenfor nævnt, at der er mange fortolkningsmuligheder, når det gælder bestemmelse af proveniensen, for er et arkivalies proveniens fx skaberen, opbevareren, brugeren eller den afleverende instans. Jeg vil mene, at de uklarheder der kan være, når det gælder papirarkivalier, ikke bliver mindre, når vi har at gøre med e-arkivalier, hvor der i en del tilfælde er tale om multi-proveniens.

Et eksempel på multiproveniens kan være Det centrale personregister (CPR) og Den ny (elektroniske) kirkebog (DNK). Sagen kompliceres her yderligere af, at afgrænsningen af e-arkivaliet eller den heuristiske enhed i forhold til IT-systemerne hos myndighederne afhænger af, på hvilket niveau systemerne betragtes. De data, der indtastes og anvendes i henholdsvis DNK og CPR, er lagret i én database. Nogle tabeller i databasen anvendes kun af CPR, enkelte tabeller anvendes kun af DNK, mens de centrale tabeller med personregistreringen anvendes af begge systemer, så på databaseniveau er der tale om én enhed. Men på brugergrænseniveau og tildels når det gælder det bagvedliggende programmel er der tale om to enheder, ligesom der er to forskellige overordnede ressortmyndigheder for de to systemer (Indenrigsministeriet og Kirkeministeriet). Det vil dog formodentlig være mest hensigtsmæssigt at betragte CPR / DNK som én heuristisk enhed for at undgå to arkiveringsversioner med en betydelig grad af redundans.

Det er Indenrigsministeriets CPR-kontor, der er ansvarlig for strukturen forstået som databasen og som programmet, der skaber funktionaliteten i CPR-systemet, mens det er Kirkeministeriet, der er ansvarlig for strukturen forstået som programmet, der skaber funktionaliteten i Den ny Kirkebog. Indholdet i CPR / DNK-databasen har derimod sin oprindelse hos en lang række (enstype) myndigheder, hvor de væsentligste er kommunernes folkeregistre, sogne, kommunale vielsesmyndigheder, domstole og statsamtter. Disse grundregistrerende myndigheder har tidligere skullet sende oplysninger til folkeregistrene, som tastede dem ind i CPR, men hensigten er, at de fremover selv skal kunne indtaste oplysninger direkte.

Et andet eksempel på multiproveniens kan være Centralt HusdyrbrugsRegister (CHR), hvor Fødevarerministeriet er skaber af og ansvarlig for strukturen, men intet har at gøre med indholdet, hverken som skaber eller bruger. Data indtastes af private aktører efter aftale med ministeriet, mens data valideres samt anvendes til sagsafgørelser i et noget kompliceret samspil mellem flere myndigheder under Fødevarerministeriet (Plantedirektoratet, Direktoratet for FødevarerErhverv og Fødevarerdirektoratet).

Der kunne nævnes flere eksempler, fx Bygnings- og Boligregistret (BBR) eller nogle af Told & Skats systemer. Og det må antages, at eksemplerne vil blive flere og flere i lyset af de bestræbelser, der er på at effektivisere den offentlige forvaltning ved bedre at udnytte data på tværs af forskellige instanser.

Pointen er, at der i forbindelse med nogle e-arkivalier er tale om multiproveniens, hvor strukturen har én proveniens, mens data har en anden eller rettere sagt mange andre provenienser. Det vil ikke være vilkårligt, hvilke data der har hvilken proveniens. Hvem der indsamler data, hvem der inddaterer, hvorfra og hvortil data overføres, samt hvem der anvender data, hænger snævert sammen med formålet og de enkelte aktørers rolle i løsningen af den eller de opgaver, som IT-systemet understøtter.

Det skal her indskydes, at begrebet proveniens i nutidig dansk arkivterminologi tildels synes erstattet af begrebet arkivskaber eller aktør:

”Der er altid mindst en arkivskaber til en heuristiske enhed..... Arkivskaber til en heuristisk enhed er den aktør eller de aktører, der enten danner den heuristiske enhed, eller systematisk nedlægger sine arkivalier i den pågældende heuristiske enhed. Formålet med registreringen af arkivskabere er ikke at skrive forvaltningshistorie, men udelukkende at understøtte arkivaliebeskrivelsen. Dette indebærer at kun aktører der har indflydelse på arkivdannelsen skal registreres.....”²⁸

Jeg vil undlade et forsøg på en fortolkning af, hvad der menes med dette, men det står tilsyneladende åbent, om arkivskaberen er skaberen af strukturen eller skaberen af indhold. Af de gældende regler for e-arkivalier fremgår det derimod klart, at der ved skaber forstås skaberen / skaberne af indhold.²⁹

Både når det gælder aktør i Dansk ArkivalieInformationsSystem, og når det gælder angivelse af skaber i arkiveringsversioner, skal der angives navne uden nærmere oplysning om den enkeltes arkivskabers rolle i forhold til struktur og (dele) af data. Det skyldes antagelig bl.a., at formålet med registreringen primært er heuristisk. Det skal som med det gamle proveniensprincip være muligt at fremsøge den heuristiske enhed eller e-arkivaliet. Ud fra det formål kan det måske være hensigtsmæssigt fx at registrere CPR-kontoret som arkivskaber til CPR-registret og ikke en næsten endeløs række af navne på samtlige kommuner, sogne, statsamter og retter.

De heuristiske hensyn er naturligvis væsentlige og kræver, at der findes ensartede og brugbare retningslinier for registrering, men mit fokus her er imidlertid ikke proveniensprincippet i heuristisk sammenhæng. Jeg vil mene, at det vil være frugtbart at anvende proveniens som et begreb, der drejer fokus mod oprindelsen af struktur og data med det formål at sikre den nødvendige forudviden for at fortolke data:

²⁸ DAISY-katekismus (Aktør), s. 1

²⁹ Skaber: Navn på alle selvstændige myndigheder, som har indtastet data i systemet. Hvis en myndigheds organisatoriske enheder ikke alle har brugt systemet eller ikke har taget systemet i brug på én gang, anføres de relevante organisatoriske enheder som arkivskabere. (Bekendtgørelse nr. 342 af 11. marts 2004 om arkiveringsversioner af bevaringsværdige data fra elektroniske arkivsystemer, bilag 4, punkt H,6)

- Hvem har skabt og haft ansvaret for strukturen (IT-systemet) og for overførslen til arkiveringsversion?
- Med hvilket formål er strukturen skabt?
- Hvem har haft ansvaret for indholdet?
- Hvorfor og hvordan er dataindsamling, datainddatering, og evt. validering foregået?
- Hvilke data er blevet overført automatisk fra andre systemer?
- Hvem har anvendt data?
- Hvordan og i hvilken kontekst er data blevet anvendt?

Ikke alle spørgsmål vil være lige relevante i forhold til forskellige e-arkivaler. Det vil fx ikke være muligt at angive navne på alle anvenderne af data i CPR / DNK, men det vil være muligt at henvise til den omfattende brug og dermed til formålet med systemerne.

Mod listen af spørgsmål ovenfor vil nogle måske indvende, at disse oplysninger ikke er nødvendige, og at det ligefrem kan tendere skrivning af forvaltningshistorie. Det kan også anføres, at hovedparten af e-arkivalier stammer fra databaser, der bevares på grund af deres informative værdi³⁰, og derfor kan deres oprindelse eller proveniens være uvæsentlig – informationerne er der jo.

Jeg er helt enig i, at mange databaser bevares på grund af deres informative værdi, og at deres oprindelse i den forbindelse ikke er væsentlig.³¹ Når man ser bort fra de databaser, der har tilknyttet elektroniske dokumenter, vil jeg mene, at ingen databaser kan bevares på grund af deres processuelle værdi.³² Den processuelle værdi findes ikke i data, men i dokumenter i tilknytning til systemet og i selve strukturen. Det kan fx være overordnede beskrivelser af formålet med systemet, sagsbehandlerinstrukser, brugervejledninger og den i systemet indbyggede funktionalitet. I nogle systemer er kriterier for fx tildeling af støtte eller beregning af skat således så eksakte, at beregningerne foretages automatisk som en del af funktionaliteten.

Jeg er derimod ikke enig i, at svarene på ovenstående spørgsmål ikke er nødvendige. Og jeg er heller ikke enig i, at det ikke er nødvendigt at sikre oplysninger om oprindelsen med henvisning til, at bevaringsbestemmelsen er truffet på grundlag af den informative værdi.

³⁰ I den Schellenbergske betydning af 'informational values' (Schellenberg, s. 6-7).

³¹ Jf. Schellenberg, s. 22: "In appraising the value of information in public records, the archivist is not greatly concerned with the source of the records The concern here is with the information that is in them."

³² I den Schellenbergske betydning af evidential values (Schellenberg, s. 6-7). Jeg finder ikke oversættelsen til processuel særlig velvalgt, men det er tilsyneladende stadig den, der er gældende i dansk arkivterminologi (Kassation og bevaring af arkivalier, s. 17-18)

For det første nåede jeg på grundlag af den teoretiske analyse med udgangspunkt i den infologiske model frem til, at viden om kommunikationsfællesskabet og dermed oprindelsen er en forudsætning for at kunne fortolke de bevarede data til information.

For det andet er oplysninger om oprindelsen af data nødvendige for at kunne vurdere pålideligheden eller reliabiliteten af dem. Med pålidelighed menes, hvor sande informationer data kan antages at repræsentere om, hvad der faktisk skete, det vil sige relationen mellem informationselementer og informationsobjekter i den infologiske model. Det kan man kun vurdere, hvis man kender dataskaberens relation til informationsobjekterne.

For det tredje er det også ud fra et historisk begrebsapparat væsentligt at have viden om oprindelsen for at kunne foretage kildekritik:

”Man bør aldrig behandle sine kilder som løsrevne stykker information, men reflektere over deres oprindelige kontekst og senere overlevering. Der er altid en grund til, at vi råder over de kilder, vi står med, og det at blive klar over oprindelse og overlevering giver ofte afgørende hjælp til at forstå kilderne og vurdere deres værdi for problemstillingen.”³³

2.4. Den infologiske model og elektroniske arkivalier

I afsnit 2.2. er fokus på fortolkningsprocessen fra data til information, som grundlæggende er den samme, uanset på hvilket medie et arkivalie er lagret. Det følgende vil derimod have fokus på særtræk ved elektroniske arkivalier, både i forhold til papirarkivalier og i forhold til IT-systemer i drift.

Jeg vil indlede med at eksemplificere ved hjælp af kirkebogen. En person er død, og denne information skal repræsenteres som et dataelement i kirkebogen. Som nævnt i afsnit 1.2. består et dataelement af seks dele eller datafragmenter, som alle er nødvendige, dog eventuelt som underforståede. Det drejer sig om objektkategori og identifikation af objektet inden for kategorien, attribut og værdien af attributen samt angivelse af tid og en specifikation af tidspunktet eller tidsintervallet.

Det kan forekomme lidt kunstigt at anvende disse begreber, men i dette tilfælde drejer det sig om objektkategorien person, og denne kan identificeres med navn og cpr-nummer. Denne person har tilknyttet attributen kirkebogshændelse, som i dette tilfælde har værdien død, og ligeledes er der angivelse af tidspunktet for denne relation mellem person og hændelse, dødsdatoen. Herudover er der i kirkebogen information, der kan betragtes som andre dataelementer, fx information om hvor, hvornår og af hvem begravelsen blev afholdt.

Den juridiske oplysning om en persons dødsfald kunne skrives som en almindelig sætning, men den er i århundreder blevet sat ind i kirkebogens faste struktur. At det drejer sig om en person, ligger implicit i selve fænomenet kirkebog, og det er forudviden i et stort kommunikationsfællesskab. At personen er karakteriseret ved at være død, fremgår af overskriften på den side, hvor han registreres samt af de fleste kolonneoverskrifter.

³³ Olden-Jørgensen, s. 50.

Dataskaberen skriver navn og cpr-nummer på afdøde i de kolonner, der ifølge overskrifterne er beregnet til dette, og ligeledes indføres dødsdato i den dertil bestemte kolonne. Det vil sige, at det dataelement, der repræsenterer information om en hændelse for en given person, dels består af datafragmenter specifikke for denne person (navn / cpr.nr og dato), dels af datafragmenter, der er indbygget i kirkebogens struktur.

Når kirkebogen afleveres til arkiv, afleveres den, som den er skabt med den udelelige sammenhæng mellem indhold og struktur.

I den nye elektroniske kirkebog vil dataskaberen skulle vælge en kirkebogshændelse svarende til at slå op i protokollen. Han vil dernæst få vist et skærbillede med felter til indtastning, og ved hvert felt vil der stå en fast ledetekst, der fortæller, hvilken oplysning (datafragment) dette felt skal indeholde. Disse faste ledetekster kan sidestilles med kolonneoverskrifterne i protokollen. På et lignende skærbillede vil der kunne slås op, hvis der er behov for at få oplysninger om afdøde personer.³⁴

For brugeren vil der således synes at være en fysisk struktur bestående af faste ledetekster og variable felter, svarende til den gamle kirkebogs kolonneoverskrifter og indskrevne oplysninger. Men synet bedrager, for det er ikke en fysisk struktur, men en præsentationsstruktur af datafragmenter samlet i dette skærbillede fra forskellige dele af IT-systemet.

De variable datafragmenter - altså de oplysninger, der varierer fra afdød til afdød som navn og dødsdato - er lagret i en relationel database. Hvis der på skærbilledet optræder et felt med alder på afdøde, kan denne oplysning være lagret i databasen, men den er det ikke nødvendigvis. Den kan i stedet fremkomme ved en beregning, der foretages hver gang, oplysningerne om en afdød vises. De datafragmenter, der er fælles for alle afdøde, er vist som ledetekster (fx 'dødsdato'), og de er defineret i applikationen, ligesom det der er defineret, i hvilke felter i den fysiske database de variable datafragmenter er lagret.

Hvis man forestiller sig brugen af Microsoft Access som det formodentlig af PC-brugere bedst kendte databaseprogrammel, kan man her ret enkelt skabe en form for applikation (formular eller rapport) på basis af ledetekster, felter med data fra databasen og evt. små beregninger, og man kan selv designe layoutet for maksimalt at understøtte brugerens fortolkning af de dataelementer, der vises på det enkelte skærbillede eller udskrift. Dette er blot som eksempel, for til systemerne bag den elektroniske kirkebog og størstedelen af de øvrige e-arkivalier er der ikke anvendt Access.

Med ovenstående eksempel har jeg forsøgt at vise, at forskellen på arkivalier på papir og i elektronisk form opstår på dataelement-niveau. På papir er datafragmenter bundet sammen til dataelementer af en struktur, der er erkendbar med det blotte øje, og som ikke kan opdeles yderligere (med mindre nogen går i gang med en saks!). Da både indhold og struktur er bundet til mediet papir overføres de uændrede til arkiv.

³⁴ Jeg har ikke haft lejlighed til at se den endelige udformning af skærbilleder i den elektroniske kirkebog, så eksemplet er tænkt og svarer muligvis ikke til den faktiske udformning.

Herimod findes et dataelement i et IT-system på skærbilleder eller i udskrifter, men dataelementet er en til et bestemt formål skabt sammensætning af datafragmenter, der stammer fra forskellige dele af IT-systemet. Hertil kommer, at selv om det enkelte dataelement på skærmen eller i udskriften fremstår som tegn, svarende til de enkelte tegn i den håndskrevne kirkebog, er det ikke den mindste betydningsadskillende enhed, idet det enkelte tegn lagres som bits, der kræver både hardware og software for at blive vist på en for mennesker tydbar vis.

Hvis e-arkivalier skal bevares svarende til den måde, hvorpå man bevarer arkivalier på papir, er den eneste mulighed at anvende den museale strategi, således at systemet og alle forudsætningerne – såvel hardware som software – for dets afvikling bevares. Den museale strategi er imidlertid opgivet som en praktisk og ressourcemæssig umulighed – det kan ikke lade sig gøre at etablere og drive et gigantisk computer-museum.³⁵

I stedet har der været overvejelser og forsøg med emuleringsstrategien, der går ud på at bevare dele af software samt dokumentation af forudsætningerne for, at software kan afvikles. Ideen er, at fremtidens computere og evt. også operativsystemer vil være mere avancerede end nutidens og derfor vil kunne bringes til at efterligne nutidens. Et af de store problemer er imidlertid, at hardware og software indgår i et kompliceret samspil, og det er svært og ressourcerelevende at få alt dokumenteret. Mig bekendt er emuleringsstrategien ikke blevet brugt i praksis nogen steder.

En tredje strategi, som anvendes i praksis, er migreringsstrategien, hvor data eller dokumenter fx flyttes fra en version af et tekstbehandlingsprogram til en ny version. Der er her altid en risiko for, at der sker utilsigtede ændringer, og da det vil være nødvendigt at foretage migrering ret ofte på grund af de hurtige skift i software, vil ændringer over tid akkumulere. Migreringsstrategien er derfor næppe brugbar i det lange perspektiv.³⁶

Den fjerde strategi er konvertering til et systemuafhængigt format, der er veldefineret, ikke-proprietært og uafhængigt af små ændringer i programmet, således at der sjældnere end ved migreringsstrategien vil være behov for konvertering til nye formater. Nogle antager dog, at der på meget langt sigt alligevel vil ske for mange datatab på grund af konverteringer. Rothenberg formulerer det på denne måde: ”This is analogous to translating Homer into modern English by way of every intervening language that has existed during the past 2.500 years.”³⁷ Et væsentligere problem med denne strategi er, at funktionalitet og struktur ikke bevares.

Med struktur menes her struktur på brugergrænsefladeniveau, men struktur kan også forstås som databasestruktur, fx relationer mellem tabeller i en relationel database. Ved design og brug af en database er det naturligvis væsentligt at sikre integriteten af data, således at data opdateres korrekt, og bl.a. i den forbindelse er relationer centrale, uanset om de er defineret på databaseniveau eller i applikationen. I arkiveringsversioner kan de give et fingerpeg om, hvilke tabeller der indeholder relaterede data, men de fortæller i

³⁵ Den kortfattede gennemgang af tekniske bevaringsstrategier er baseret på Rothenberg.

³⁶ Begreberne migrering og konvertering anvendes tilsyneladende lidt i flæng. Migrering synes at være en undertype af konvertering, hvor data flyttes inden for fx forskellige versioner af samme software., ikke til fx et systemuafhængigt format.

³⁷ Rothenberg, s. 11.

de fleste tilfælde ikke, hvilke felter fra de relaterede tabeller, der tilsammen skaber et dataelement. I ICA's Guide defineres de to typer af struktur således:

"Database structures, in addition to the above logical structure, also include presentation structures. Whereas the logical structure organize data for storage and processing, the presentation structure are visible structures used by individuals to input data into the database and to access and view data."³⁸

Uanset hvilken strategi der anvendes, vil der ske en splittelse af medie, indhold og struktur, når IT-systemer ændres til e-arkivalier. Med udgangspunkt i den infologiske model kan der peges på en række områder, hvor der er behov for bevaring eller bevaring af dokumentation. Listen nedenfor er rangordnet, idet opfyldelsen af punkt et er en forudsætning for, at opfyldelsen af punkt to har nogen mening og så videre:

- Bevaring af de mindste betydningsadskillende enheder, bits, ved at sikre, at lagringsmedier ikke forgår, og ved at sikre dokumentation af, med hvilket udstyr de skal læses.
- Bevaring af mulighed for omformningen af bits til de for mennesket mindste betydningsadskillende enheder, tegn, ved at sikre, at der findes programmer eller tilstrækkelig dokumentation til at kunne skabe programmer, der kan foretage omformningen.
- Bevaring af de datafragmenter (felter), der findes i det oprindelige IT-systems database, samt dokumentation af informationen i det enkelte datafragment.
- Bevaring af dokumentation af de dele af applikationen, der har skabt dataelementerne ved på brugergrænsefladeniveau at definere struktur.
- Bevaring af dokumentation af de dele af software, der har haft betydning for data i databasen. Det vil sige de beregninger, der foretages automatisk, og de dataoverførsler der sker fra andre systemer.

Der skal her indskydes, at ovenstående udredninger drejer sig om databaser, ikke elektroniske dokumenter, der er lagret som TIFF i elektroniske sags- og dokumenthåndteringssystemer. Den mindste betydningsadskillende enhed er også her bits, hvor der kræves programmel for at omsætte disse til et billede på skærmen. Men derefter har man et dokument, der svarer til et papirdokument, uden problemer med at identificere dataelementer. Elektroniske dokumenter lagret som TIFF vil naturligvis blive betragtet som e-arkivalier, men de er et særtilfælde. Resten af det elektroniske sags- og dokumenthåndteringssystem, det vil sige de registreringer, der er nødvendige for at finde og sammenkæde dokumenterne til sager, er kendetegnet ved den samme fragmentering af dataelementerne som beskrevet ovenfor i eksemplet med kirkebogen.

³⁸ Guide for Managing Electronic Records, del 2, afsnit 1.1

2.5. Autenticitet

At formålet med et arkiv er at bevare arkivalier, der kan tjene som autentisk dokumentation eller autentiske kilder, står næppe til diskussion. Det er så indlysende, at begreberne autentisk og autenticitet ikke behøver forklaring - jeg har i hvert fald haft svært ved at finde definitioner. Men hvad menes der egentlig med begrebet autenticitet i arkivsammenhæng, og kan e-arkivalier være autentiske? Jeg vil nedenfor pege på nogle svar, men de vil ikke være fuldstændige. Det er et område, der kunne trænge til nærmere undersøgelse og diskussion.

Begrebet har selvfølgelig en del at gøre med den dagligdags, positivt ladede betydning af ordet: ægte, oprindelig, uforfalsket, som det så ud fra ophavsmandens side (af græsk *authentēs*: ophavsmand). Det er som sagt svært at finde en arkiv-definition, men i Guiden forklares og afgrænses begrebet i arkivsammenhæng:

“The reliability of a record is its ability to serve as reliable evidence. Basically, a record can be no more reliable than it was at the instant of its creation. Therefore, direct responsibility for reliable records is that of the records creator.... Authenticity refers to the persistence over time of the original characteristics of the record with respect to context, structure and content. An authentic record is one that retains its original reliability.”³⁹

I denne forklaring anvendes ud over begrebet autenticitet begrebet pålidelighed, og pålideligheden er knyttet til arkivaliets potentiale for at tjene som vidnesbyrd. I infologiske termer kan det beskrives sådan, at arkivaliet er pålideligt, hvis dataelementerne i arkivaliet repræsenterer informationselementer, der fortæller noget sandt om informationsobjekterne. I historiske termer kan det formuleres således, at arkivaliet er pålideligt, hvis ophavsmanden og ophavssituationen tilsiger, at oplysningerne i arkivaliet fortæller noget, der må anses for at være i overensstemmelse med, hvad der faktisk skete. Jeg vil dog ikke komme nærmere ind på dette. For det første fordi det er en teoretisk kompliceret problematik, og for det andet fordi det slet ikke vedkommer arkivet! Et arkiv skal ikke blande sig i ophavssituationen – højst sikre dokumentationen af den, men arkivet har forpligtelsen til at bevare data autentisk. Heraf følger, at data, der er fejlagtige, er autentiske, hvis de forbliver fejlagtige!

Autenticitet er ifølge forklaringen ovenfor, at arkivaliet bevarer sine originale karakteristika, hvad angår kontekst, indhold og struktur, og at det bevarer den oprindelige grad af pålidelighed som vidnesbyrd. Når det gælder papirarkivalier, vil jeg hævde, at de bevarer deres autenticitet, vel vidende at det kan problematiseres. Er et dokument fx autentisk, selvom det er en fotokopi eller er mikrofilmet eller scannet? Der er sket en overflytning til et andet medie, men sammenhængen mellem indhold og struktur bevares, hvad der for mig at se er det væsentligste.

E-arkivalier derimod bevarer ikke deres originale karakteristika, med mindre den museale strategi anvendes, og det er som sagt urealistisk. Uanset om den valgte strategi baseres på emulering, migrering eller konvertering til systemuafhængigt format, sker der en splittelse af medie, indhold og struktur. Bits vil blive flyttet fra et medie til et andet.

³⁹ Guide for Managing Electronic Records, afsnit 2.1.

Indholdet synes forhåbentlig uforandret, men på bits-niveau kan der ske ændringer fx ved tegnsætskonvertering. Strukturen vil ligeledes blive ændret eller falde bort, når konvertering til systemuafhængigt format anvendes. Så konklusionen er, at elektroniske arkivalier ikke kan være autentiske!⁴⁰ Det skal indskydes, at dette ikke blot gælder databaser, men også elektronisk lagrede dokumenter, hvilket er et område, der er fokus på i flere internationale projekter.⁴¹

Når e-arkivalier ikke kan bevares som autentiske i traditionel forstand, skal det på anden vis sikres, at deres autenticitet kan påvises. Det er derfor nødvendigt at dokumentere, hvilke ændringer der sker. For det første for at arkivet har et grundlag for at garantere arkivaliets autenticitet. For det andet således at der i fremtiden vil være en mulighed for så at sige at gå baglæns i de omformningsprocesser, der er sket, indtil man når tilbage til det oprindelige IT-system. Det foreslås undertiden, at konverterings- og migreringsprocesser skal være reversible, men det vil nok være problematisk, for den oprindelige hard- og software må antages at være borte. Så processerne kan kun være reversible i simuleret form.

Behovet for bevaring af autenticiteten gennem dokumentation af omformningsprocesser opstår ikke først, når IT-systemet bliver til et arkivalie, der afleveres til arkiv. De omformningsprocesser, der sker hos arkivskaberen, fra data lagres, og til de overføres til arkiveringsversion, bør også dokumenteres, om end det kan diskuteres, om det er noget, et arkiv skal blande sig i. Det kan fx være konvertering til en anden database, nye skærbilleder eller migrering af dokumenter til et andet tekstbehandlingssystem.

Som sagt er der behov for, at problematikken om autenticitet undersøges og diskuteres nærmere. Det er i den forbindelse væsentligt at få afklaret, hvad der forstås ved autenticitet, og hvor høje krav arkivet og fremtidige forskere må antages at stille til den dokumentation, der skal være grundlaget for at påvise, at et e-arkivalie er autentisk på trods af omformninger.

Med udgangspunkt i den infologiske model kan jeg pege på nedenstående områder, hvor der kan være behov for dokumentation gennem hele IT-systemets og arkiveringsversionens livscyklus:

- Omformninger på bit-niveau. Når bits populært sagt flyttes fra et lagringsmedie til et andet, skal det dokumenteres, at det er kontrolleret, at der ikke er sket ændringer. Hvis bits ændres på grund af konvertering til andet tegnsæt, bør enten hver enkelt ændring eller type af ændring dokumenteres. Dette vil også gælde, hvis der sker konvertering af elektroniske dokumenter.
- Omformninger på datafragment-niveau, det vil fx sige, hvis der sker ændringer i databasestrukturen. Sådanne omformninger kan både ske, mens systemet er i drift, ved konvertering fra IT-system til arkiveringsversion og ved konvertering hos arkivet.

⁴⁰ ”In the case of electronic texts the useful will be more important than the authentic.” (Edward Higgs i *The Concept of Record*, s. 110).

⁴¹ Fx i INTERPARES-projektet. Det diskuteres ligeledes i Nielsen.

- Ændringer af struktur på brugergrænsefladeniveau. Hvis strukturen helt falder bort, fordi data afleveres systemuafhængigt, må strukturen dokumenteres på anden vis.

2.6. Opsummering

Jeg er kommet med nogle svar på de tre indledende spørgsmål om forskelle på arkivalier og anden information, forskelle på papirarkivalier og e-arkivalier og forskelle på IT-systemer i drift og e-arkivalier. Det har jeg gjort med udgangspunkt i infologien, men jeg har også forsøgt at inddrage arkiv-begreberne arkivalie, proveniens og autenticitet. Undervejs har jeg peget på forskellige behov for dokumentation med det mål at få et grundlag for at opstille den 'ideelle' liste over, hvad der skal dokumenteres, når konvertering til systemuafhængig arkiveringsversion anvendes.

Arkivalier indeholder data, der repræsenterer information, men den karakteristik gælder også for bøger, tidsskrifter og så videre. Arkivalier adskiller sig ved at indeholde data, der er skabt med henblik på fortolkning til information i en begrænset, veldefineret kontekst og ikke i en kontekst, der er udstrakt i både tid og rum. Når data ikke længere skal anvendes i den oprindelige kontekst, overføres de til arkiv på grund af genanvendelsesværdien, og samtidig løsrives de fra konteksten. Konteksten kan imidlertid delvist dokumenteres, og firkantet kan det siges, at jo mere dokumentation eller jo flere kilder, der er til konteksten, jo bedre grundlag har en fremtidig forsker for at fortolke data til information.

Når papirarkivalier afleveres til arkiv, bevares sammenhængen mellem medie, indhold og struktur, da de er fysisk uadskillelige. Sammenhængen kan derimod ikke bevares, når IT-systemer bliver til e-arkivalier. Sammenhængen og strukturen må i stedet dokumenteres. Ligeledes må ændringer fra IT-system til arkiveringsversion og konvertering efter aflevering til arkiv dokumenteres i det omfang, det er nødvendigt for at kunne påvise, at e-arkivalierne er autentiske trods ændringer.

I hovedafsnit 4 vil jeg vende tilbage til behovene for dokumentation og komme med forslag til den 'ideelle' liste. Med henblik på at få inspiration til denne har jeg imidlertid undersøgt, om der internationalt er standarder eller projekter, der fokuserer på det samme. I det følgende hovedafsnit 3 vil jeg redegøre for nogle af de projekter og standarder, som jeg har vurderet som mest vægtige, både på grund af deres indhold og på grund af den hyppighed, hvormed der henvises til dem fra andre projekter og fra artikler. Jeg må dog sige, at selv om de refererede projekter og standarder er interessante og tankevækkende, er det begrænset, i hvor høj grad de har kunnet give svar i forhold til min specifikke problematik.

3. INTERNATIONALE PROJEKTER OG ANBEFALINGER

I forbindelse med dette projekt har jeg naturligvis søgt efter litteratur med fokus på e-arkivalier, og der findes en del materiale i form af artikler, rapporter samt hjemmesider for forskellige projekter og organisationer. En del er som sagt præget af en vis forvirring og går ikke rigtig i dybden med problemstillingerne, og det er ofte den samme lille personkreds, der ytrer sig. Andet er fokuseret på de særdeles væsentlige tekniske problemer med bevaring af bits og data, mens der er mindre fokus på bevaring af information. Der er ikke meget om konkrete erfaringer med bevaring og endnu mindre om erfaringer med brug af e-arkivalier, altså fortolkningen af data til information efter aflevering til arkiv, af den simple grund at det stadig er begrænset, hvor mange e-arkivalier der er afleveret til arkiver.

Men jeg har udvalgt nogle projekter og anbefalinger til en nærmere beskrivelse nedenfor. Udvælgelsen er sket ud fra tre kriterier: En vis relevans for dette projekt, omfang og kvalitet forstået på den måde at de ikke virker overfladiske, samt hyppighed af henvisninger til dem.

3.1. Open Archival Information System (OAIS)

I Open Archival Information System skal system ikke forstås som IT-system, men som organisation (OAIS anvendes synonymt med arkiv), og åben skal ikke forstås som et arkiv uden restriktioner for adgang, men som et arkiv der overholder OAIS-anbefalingerne, der er udviklet i et åbent miljø. Anbefalingerne er udviklet på opfordring fra bl.a. ISO, uden at de dog på nuværende tidspunkt er en ISO-standard, og bag udviklingen står først og fremmest den internationale rumfartsorganisation.

Formålet er at definere rammer og begreber for, hvilke funktioner et arkiv skal varetage for at kunne sikre langsigtet bevaring af information, og der er opstillet en model for organiseringen af arkivet, ligesom der er defineret en lang række begreber for funktioner og produkter. OAIS ser ud til det sidste par år at være blevet et udbredt begreb i arkivverdenen, ligesom nogle arkiver overvejer at anvende modellen, fx har det svenske Rigsarkiv oprettet et projekt, der sigter mod et OAIS-arkiv.

OAIS-anbefalingerne drejer sig derimod ikke om konkrete forslag til bevaringsstrategi, men der er dog nogle eksempler, som tyder på, at dokumenter og data enten forventes arkiveret systemuafhængigt eller i det oprindelige format, hvilket medfører overvejelser om, at dette på sigt vil kunne kræve konvertering. Emulering nævnes også som en mulighed, hvis 'look and feel' skal bevares, men der er mange forbehold over for denne strategi.⁴²

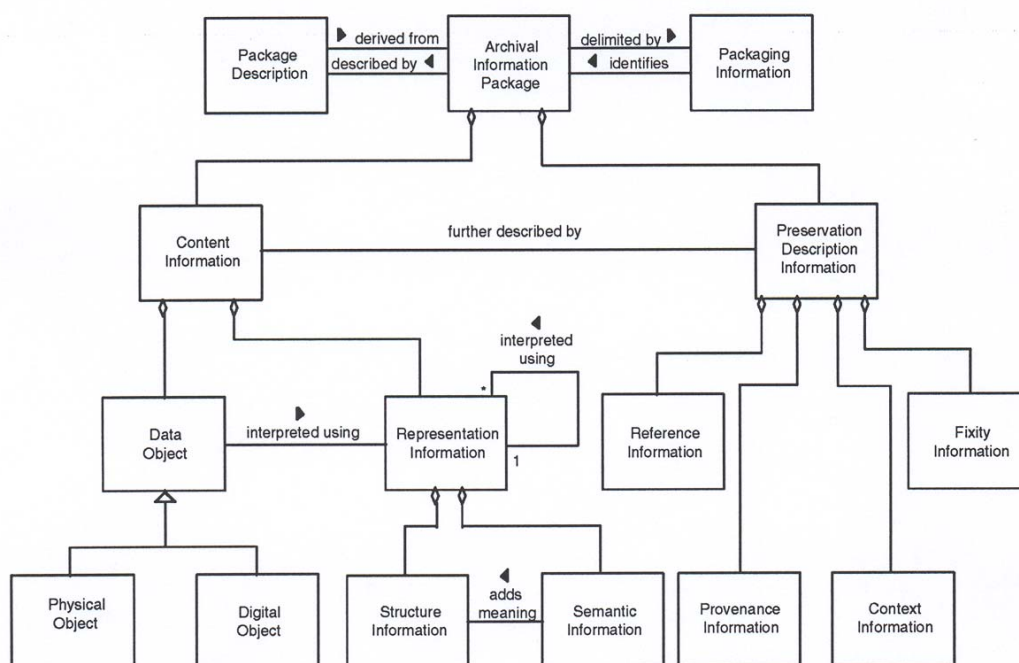
I anbefalingerne nævnes forskellen på data og information og ligeledes behovet for en 'knowledgebase' for at kunne fortolke data til information.⁴³ Men eksemplerne på 'knowledgebase' drejer sig om sprog og om fagudtryk, ikke om en mere generel forudviden. Det skyldes antagelig, at anbefalingerne retter sig mod arkiver, der arkiverer data

⁴² Reference Model for an Open Archival Information System, s. 5-10 til 5-12

⁴³ Reference Model for an Open Archival Information System, s. 2-3

rettet mod et 'designated community', hvilket defineres som "an identified group of potential consumers who should be able to understand a particular set of information"⁴⁴. Det synes således at være forudsat, at brugerne i forvejen tilhører kommunikationsfællesskabet, hvilket ikke kan undre i betragtning af oprindelsen af OAIS i rumfartsorganisationer. Samtidig antages det, at data kontinuerligt vil blive brugt, så fx begreber løbende kan omdefineres, når kommunikationsfællesskabet ændrer sig over tid.

Det centrale produkt i et OAIS-arkiv defineres som Archival Information Package (AIP), og det omfatter selve data samt den for fortolkningen nødvendige dokumentation. Med danske ord kunne AIP betegnes som en arkiveringsversion.



Figur 4⁴⁵

Som det fremgår af figur 4 består en AIP af en række elementer, der kort kan defineres på nedenstående vis.

Package Description og *Packaging Information* indeholder oplysninger, der skal muliggøre fremfindning af filerne.

Content Information defineres som det sæt af informationer, der er det oprindelige mål for bevaring. *Content Information* opdeles for det første i *Data Object*, der er selve data forstået som bitstrømmen. For det andet består *Content Information* af *Representation Information*, der er de oplysninger, som er nødvendige for at kunne få meningsfuld information ud af bitstrømmen.

Representation Information består for det første af *Structure Information*, der drejer sig om organiseringen af data, hvor der som eksempel nævnes de oplysninger, der er nød-

⁴⁴ Reference Model for an Open Archival Information System, s. 1-10

⁴⁵ Reference Model for an Open Archival Information System, s. 4-37

vendige, for at bitstrømmen kan omformes til tegn ved hjælp af ASCII-standarden. For det andet består Representation Information af *Semantic Information*, der er oplysninger på det sproglige niveau, fx at sproget er engelsk og forklaringer til fagudtryk. Ved Representation Information er der markeret en rekursiv funktion, som skal vise, at en AIP ikke nødvendigvis er selvdokumenterende, men kan indeholde henvisninger til Representation Information uden for AIP, fx når det gælder tegnsætsdefinitionen, som man kan nøjes med at have defineret et sted og ikke i alle AIP'er, hvor netop det tegnsæt anvendes.

Preservation Description Information defineres som de informationer, der er nødvendige for en fyldestgørende bevaring af Content Information, og den består af fire dele. *Reference Information* er oplysninger om identifikatorer, hvor ISBN nævnes som eksempel. *Fixity Information* er oplysninger om mekanismer og nøgler til sikring af autenticiteten. *Provenance Information* defineres som dokumentationen af historien bag Content Information. Det er oplysninger om oprindelsen eller kilden til data, hvilke ændringer data har undergået, siden de blev skabt, og oplysninger om hvem der over tid har haft ansvaret for data. *Context Information* defineres som dokumentation af relationen mellem data og deres omgivelser, fx hvorfor data er skabt, og hvordan de forholder sig til andre forekomster af Content Information.

Det må konkluderes, at OAIS ikke direkte giver svar på mine spørgsmål om, hvordan fortolkning af data til information kan sikres bedst muligt, når bevaring sker ved hjælp af den systemuafhængige konverteringsstrategi. I sammenhæng med min problematik vil jeg pege på følgende som mest interessant og tankevækkende:

- At funktionerne i arkivet betragtes som ét system, således at funktionerne og produkterne fra modtagelse over bevaring til adgang tænkes som en sammenhængende helhed.
- At data ikke betragtes som løsrevet eller overordnet, men er sidestillet med Representation Information. I forhold til den infologiske model vil jeg mene, at Representation Information indeholder de oplysninger, der er nødvendige for omdannelsen fra bit til tegn, fra tegn til datafragment og fra datafragment til dataelement.
- At det kan være vanskeligt at gøre en AIP (arkiveringsversion) fuldt ud selvforklarende, og at det kan være en ide at have henvisninger til Representation Information, der er arkiveret andetsteds.

3.2. InterPARES

The International Research on Permanent Authentic Records in Electronic Systems, kaldet InterPARES, er funderet i Canada ved University of British Columbia, men har deltagere fra arkiver i en række lande, bl.a. Australien, Kina, Holland og Italien. InterPARES1 arbejdede i perioden 1999 – 2001 med spørgsmål vedrørende langsigtet bevaring af autentiske elektroniske arkivalier, hvor arkivalier som udgangspunkt forstås som dokumenter. I 2002 påbegyndtes InterPARES2, hvor hensigten er at arbejde videre med spørgsmål vedrørende autenticitet gennem hele et elektronisk dokumentets livscyklus samt med autenticiteten af andre typer arkivalier end dokumenter, fx lyd og billeder. På

nuværende tidspunkt kan der imidlertid kun findes en kortfattet præsentation af InterPARES2, der planlægges at være til 2006.

Derimod er der offentliggjort et omfattende materiale om InterPARES1.⁴⁶ Projektet var delt i tre delprojekter (kaldet Task Forces), der arbejdede henholdsvis med en overordnet begrebsmæssig afklaring af, hvad der konstituerer et elektronisk arkivalie og dets autenticitet, med spørgsmål vedrørende bevaring og kassation samt med problematikker vedrørende bevaring af autenticitet efter modtagelse i arkivet. Desuden blev der på basis af nogle af projekternes erfaringer opstillet en liste over meget generelle punkter, som bør tages i betragtning, når elektroniske arkivalier skal bevares.⁴⁷

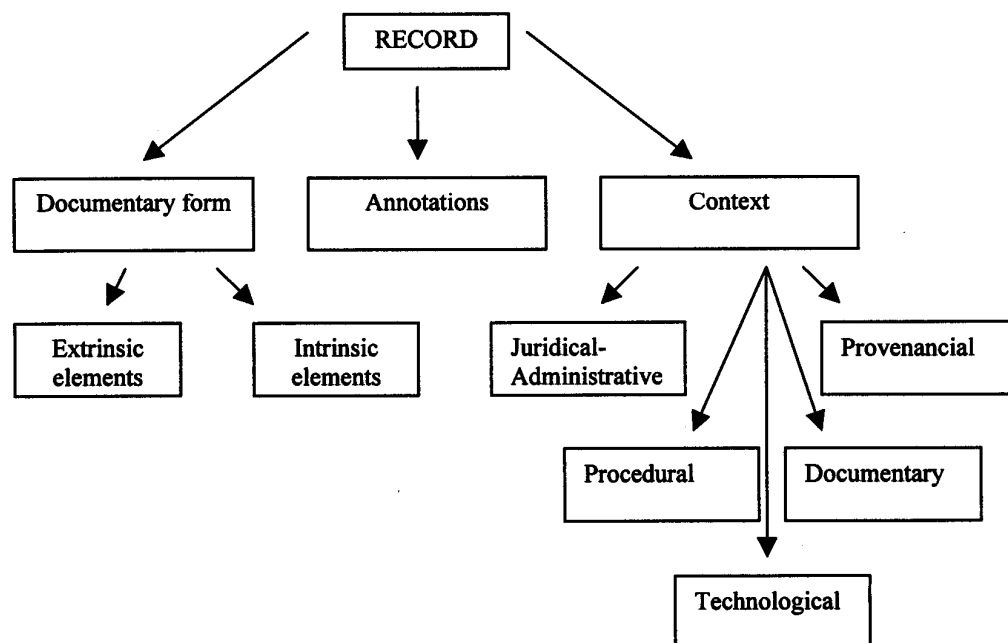
Projekterne i INTERPARES1 betegnes som forskningsprojekter, hvis sigte har været at komme med konkrete svar på en række spørgsmål, og der er også en række anbefalinger i de endelige rapporter, men samtidig rejses der mange nye spørgsmål. Der lægges heller ikke skjul på, at selve processen har været krævende, fordi deltagernes baggrund var vidt forskellig både geografisk og fagligt, og bare det at nå til en nogenlunde fælles forståelse af begreber var meget tidskrævende. Efter min mening indeholder projektrapporterne dog en udvikling i forhold til den spredte fægtning, der afspejles i de internationale diskussioner fra midten af 1990'erne, både hvad angår en begrebsmæssig afklaring og en praktisk orienteret vurdering af de opgaver, som elektroniske arkivalier stiller arkiverne over for.

*Taskforcen vedrørende autenticitet*⁴⁸ havde bl.a. som mål at definere, hvilke fællestræk eller elementer der karakteriserer elektroniske arkivalier og er med til at konstituere autenticiteten, samt at afklare, om elementerne kan variere afhængig af typen af elektronisk arkivalie. Det teoretiske udgangspunkt er diplomatikken, på basis af hvilken der blev opstillet en skabelon, som jeg har forsøgt at illustrere med nedenstående figur 5.

⁴⁶ The Long Term Preservation of Authentic Electronic Records: Findings of the InterPARES Project (www.interpares.org)

⁴⁷ Strategy Task Force Report i The Long Term Preservation of Authentic Electronic Records

⁴⁸ Authenticity Task Force Report i The Long Term Preservation of Authentic Electronic Records



Figur 5⁴⁹

Documentary form defineres som ”the rules of representation according to which the content of a record, its administrative and documentary context, and its authority are communicated.”⁵⁰ Den dokumentariske form deles i ydre og indre elementer, hvor de ydre elementer groft sagt er knyttet til udseendet af et dokument, mens de indre elementer er knyttet til indholdet. De ydre elementer kan fx være layout, vandmærke, logo, elektronisk signatur, tidsstempling. De indre elementer er de oplysninger i indholdet, der er med til at bekræfte autenticiteten, fx navne og datoer.

Annotations defineres som ”additions made to a record after it has been created as part of the formal execution phase of an administrative procedure.”⁵¹ Det kan fx være oplysninger om afsendelse, modtagelse og journalisering.

Context defineres kortfattet som ”the framework of action in which the records participates”.⁵² Konteksten består for det første af juridisk-administrativ kontekst, der er det juridiske og administrative system, som arkivskaberen tilhører. For det andet består den af proveniens-konteksten, der drejer sig om arkivskaberen og dennes struktur, mandater og funktioner. For det tredje består den af procedure-konteksten, og for det fjerde af den dokumentariske kontekst, der drejer sig om arkivaliets indre struktur og dets placering i arkivsystemet. For det femte består konteksten af den teknologiske kontekst, der defineres som de tekniske komponenter i det elektroniske system, hvori arkivaliet er skabt. I

⁴⁹ Figur baseret på Template for Analyses, Appendix 1 i The Long Term Preservation of Authentic Electronic Records

⁵⁰ Ibid, s. 1

⁵¹ Ibid, s. 4

⁵² Ibid, s. 5

denne kontekst indgår både hardware, software, data forstået som filstruktur og data-formater, systemmodeller forstået som beskrivelser af såvel databaser som funktionalitet samt systemadministration.

Som sagt er skabelonen over, hvilke elementer der er med til at konstituere et arkivalies autenticitet, lavet som et teoretisk arbejde baseret på diplomatikken. Taskforcens erfaringer med at anvende skabelonen på faktiske arkivalier var imidlertid problematik, for de levede ikke op til teorien! Først forsøgte man med arkivalier, der var inaktive, men det viste sig generelt, at der var alt for lidt bevaret af konteksten. Så forsøgte man med case-studies af arkivalier i aktive systemer, men det var for det første svært at afgrænse, hvad der var arkivalier i systemerne, fordi systemer var komplicerede i deres blanding af data og funktionalitet, blanding af database og dokumenter samt sammenkædningerne mellem flere databaser. Desuden havde man haft den forventning, at jo mere normativt et dokument er, jo flere og bedre er elementerne, der konstituerer dets autenticitet. Det viste sig ikke at være tilfældet.

*Task Forcen vedrørende bevaring og kassation*⁵³ diskuterer ikke kriterier for bevaring, men fokuserer på, hvornår og efter hvilken fremgangsmåde vurderingen skal finde sted. Uanset hvilken fremgangsmåde et arkiv tidligere har anvendt i forhold til arkivalier på papir, er det en nødvendighed at foretage bk-vurdering længe før aflevering, evt. før systemet er taget i brug. Vurderingen skal som altid foretages på grundlag af den langsigtede historiske og dokumentariske værdi, men også ud fra undersøgelser af hvor godt autenticiteten er sikret i de kørende systemer, og ud fra analyser af hvor omkostningsfyldt en langsigtet bevaring vil være i forhold til værdien af data, altså en form for cost-benefit vurdering. For at foretage analyser og vurdering skal der indsamles et omfattende materiale, og dette materiale er centralt for den fremtidige forståelse og brug af arkivalier, og det skal derfor også bevares. Ud over bevaringsvurderingen skal arbejdet også resultere i præcise retningslinier for, hvordan data på et senere tidspunkt skal overføres til arkiv. Indtil data overføres til arkiv, skal både bevaringsbeslutning og bestemmelser om aflevering overvåges, da al erfaring siger, at systemer ændrer sig over tid, ofte over særdeles kort tid, således at der kan være behov for justeringer i bestemmelserne.

Taskforcen vedrørende bevaring og kassation anvender implicit definitionen af en record som et dokument, og i forbindelse med en konkret case fører det til fokus på de udfordringer, som elektroniske sagsbehandlingssystemer kan medføre. Der er tale om et patentsystem, der består af en relationel database med et større antal tabeller og en række standardbreve, der flettes med data fra databasen inden udskrivning. Brevene fremkommer således på basis af forskellige komponenter fra systemet, og de gemmes ikke i systemet, da processen kan gentages, så længe systemet er i drift. Også denne problematik skal der tages stilling til ved bevaringsvurderingen, for "in order for the archives to preserve authentic electronic records over time, the appraiser must be aware of what these components are, what elements are contained therein, and the means by which the elements can be united to reproduce the record in a comprehensive form."⁵⁴

⁵³ Appraisal Task Force Report i The Long Term Preservation of Authentic Electronic Records

⁵⁴ Appraisal Task Force Report, s. 12 (i The Long Term Preservation of Authentic Electronic Records)

*Task Forcen vedrørende bevaring*⁵⁵ af først og fremmest autenticitet efter aflevering til arkiv når frem til, at det kan der ikke opstilles generelt gældende tekniske løsninger for, fordi elektroniske arkivalier er meget forskellige, og fordi bevaring af autenticitet afhænger af, hvordan den blev sikret, da det givne arkivalie blev skabt og anvendt, det vil sige arkivaliets kontekst, inden det blev arkivalie.⁵⁶ For ethvert elektronisk arkivalie gælder det dog, at det ikke kan bevares – det er kun muligheden for at reproducere det, der kan bevares. Input til bevaringsprocessen defineres som de arkivalier, der skal bevares, samt de nødvendige informationer for at bevare dem autentiske og tilgængelige. Adgang til arkivalierne samtænkes imidlertid i en sådan grad med bevaring, at input til bevaringsprocessen også er forespørgsler om adgang til arkivalierne eller oplysninger om dem. Output'et af bevaringsprocessen er derfor reproducerede arkivalier (arkivet har reproducet dem) eller reproducerbare arkivalier (data er ledsaget af de nødvendige oplysninger, så brugeren selv kan reproducere arkivalierne) samt evt. et certifikat, der garanterer autenticiteten.

Sammenfattende kan siges, at InterPARES i sit udgangspunkt er begrænset til fokus på bevaring af autenticitet i elektroniske dokumenter, men spørgsmålet om databaser berøres indirekte, bl.a. fordi dokumenter i en elektronisk verden ikke er et entydigt begreb, jf. eksemplet med patentsystemet. Af anbefalinger fra projekterne vil jeg især pege på følgende:

- Skabelonen over hvilke elementer der er med til at sikre et arkivalies autenticitet, idet denne skabelon har en noget anderledes indgangsvinkel end fx OAIS-anbefalingerne til, hvad der skal bevares udover selve data. Der er dog også mange sammenfald, hvor forskellene tildels er forskellige placering og navngivning af de enkelte elementer.
- At bevaring- og kassationsvurdering skal ske på et tidligt tidspunkt, gerne før et system tages i brug, og at vurderingen skal baseres på en grundig analyse, som også munder ud i detaljerede retningslinier for fremtidig arkivering. Såvel disse retningslinier som bevaringsvurderingen skal løbende revurderes, når systemet ændres. Det arkivskabte materiale, der fremkommer ved dette arbejde, er væsentligt at bevare.
- At e-arkivalier ikke kan bevares som papirarkivalier, men at fokus skal være på muligheden for at reproducere de elektroniske arkivalier sammen med dokumentation af de elementer, der konstituerer deres autenticitet. Projektet peger på, at det i arkivverdenen har været svært at erkende, at elektroniske arkivalier ikke kan bevares på parallel vis til arkivalier på papir.

3.3. International Council on Archives: Guiden 1997 og Håndbogen 2004

I midten af 1990'erne udgav ICA som tidligere nævnt en oversigt over en stor del af den engelsksprogede litteratur, der på daværende tidspunkt fandtes om den langsigtede be-

⁵⁵ Preservation Task Force Report i *The Long Term Preservation of Authentic Electronic Records*

⁵⁶ Denne skellen mellem data i brug og arkivaliet anvendes dog ikke så eksplicit i InterPARES-projektet, hvor livscyklus-tankegangen til dels ligger som baggrund. InterPARES-projektets forgænger fra midt i 1990'erne, det såkaldte UBC-projekt (University of British Columbia), havde således fokus på sikring af autenticitet hos records i brug.

varing af elektroniske arkivalier, primært artikler og indlæg på konferencer. Der er mange interessante overvejelser og diskussioner, men det generelle indtryk er, at meget af litteraturen er meget abstrakt, at diskussionerne så at sige stritter i alle retninger, og at man taler forbi hinanden. Omtrent samtidig udgav ICA en guide vedrørende håndteringen af elektroniske arkivalier. Denne guide indeholder de første forsøg på at komme med konkrete anvisninger på, hvordan et arkiv skal håndtere de nye udfordringer.

Med udgangspunkt i denne guide er der blevet arbejdet på en håndbog i elektroniske arkivalier for arkivarer. Håndbogen blev præsenteret og diskuteret på ICA-kongressen i Wien og forventes publiceret i slutningen af 2004. Formålet med den beskrives at være at give arkiverne forslag til, hvordan deres politik kan reorienteres, så de kan varetage elektroniske arkivalier. Forslagene skal omfatte alle typer af arkivalier og skal kunne bruges, uanset om de bevares hos skaberen eller overføres til arkiv.

Håndbogen skal således omfatte meget, men den er overskueligt bygget op, og jeg vil betegne den som en checkliste over, hvilke opgaver et arkiv bør overveje at varetage i forhold til elektroniske arkivalier, samtidig med at håndbogen på et overordnet niveau kommer med forslag til, hvordan opgaverne kan løses. I det følgende giver jeg et meget kort resume af håndbogens indhold:

To afsnit er viet spørgsmål om, hvordan arkivdannelsen kan påvirkes, eller måske rettere sagt hvordan man kan forsøge at sikre, at arkivmæssige hensyn varetages, så der faktisk bliver dannet arkiver. Det ene afsnit behandler spørgsmålet på det overordnede strategiske niveau og det andet spørgsmålet i forhold til de enkelte systemer.

Et afsnit drejer sig om den langsigtede bevaring, hvormed menes bevaring udover levetiden for det system, som et arkivalie er skabt i. Anbefalingen er her en eller anden form for systemuafhængig bevaring i et veldokumenteret og åbent format. Det bliver her tydeligt, at såvel den museale strategi som emuleringsstrategien er næsten opgivet og kun anses for brugbare i særlige tilfælde. Det nævnes imidlertid, at selv om data bevares systemuafhængigt, kan man forsøge at bevare muligheden for at genindlæse dem i det system, som de oprindeligt blev skabt i. For mig at se vil det bortset fra det helt korte sigt indirekte kræve en museal strategi.

Vigtigheden af metadata betones, og der defineres tre forskellige typer af metadata. For det første såkaldte record-keeping metadata, der er de metadata, der skabes sammen med selve arkivaliet eller dokumentet, fx journaliseringsoplysninger. For det andet arkiv-metadata, der af arkivskaberen eller arkivet tilføjes et arkivalie, efter det er gået ud af primær brug, fx proveniensoplysninger. For det tredje tekniske metadata, der skal gøre det muligt at forstå og anvende data. Det er en meget grov inddeling af metadata, men dog et fremskridt i forhold til de mange tilfælde, hvor metadata nævnes som centrale for bevaringen, men hvor det ikke præciseres, hvad der menes, bortset fra at metadata fortæller et eller andet om de data, der skal bevares.

Det sidste ret kortfattede afsnit drejer sig om acces, og det er præget af, at der er gjort færre praktiske erfaringer med denne opgave end med de ovenfor nævnte. Det understreges, at der er et utal af løsninger, at ingen løsning er bedre end andre, og at valget er det enkeltes arkivs. Ved valget af løsning bør arkivet afbalancere brugerbehov, arkivets

eget behov for serviceniveau udadtil samt omkostninger. Herefter kan man se på, hvordan acces teknisk skal løses.

3.4. ISAD(G): General International Standard Archival Description

ISAD(G) er den internationale, af ICA vedtagne standard for beskrivelse af arkiver og arkivalier. Med hensyn til beskrivelsen af proveniens eller måske rettere arkivskabere knytter ISAD(G) an til ISAAR(CPF): International Standard Archival Authority Record for Corporate Bodies, Persons and Families. ISAAR(CPF) vil jeg ikke beskæftige mig yderligere med, da denne standard som sagt drejer sig om beskrivelse af arkivskabere og ikke om beskrivelse af arkivalier.

Til ISAD(G) er der udarbejdet et SGML/XML-opmærkningsbibliotek, kaldet Encoded Archival Description (EAD). Der ser ligeledes ud til at være et arbejde i gang med at lave et tilsvarende opmærkningsbibliotek for ISAAR(CPF), kaldet Encoded Archival Context (EAC).

Den første version af ISAD(G) blev publiceret i 1994 og kom i en revideret anden version i 2000. Standarden indeholder 26 såkaldte elementer, der er placeret under syv områder: identity statement area, context area, content and structure area, conditions of access and use area, allied material area, notes area, description control area. Beskrivelsen med de 26 elementer kan være på fire forskellige niveauer, idet arkivalierne ses som hierarkisk opbygget. De fire niveauer er fond, serie, file og item. Item på det nederste niveau er den mindste enhed, fx et dokument. File er en samling af dokumenter, der er samlet af skaberen eller arkivet, fordi de drejer sig om det samme emne eller er resultatet af en aktivitet. Serie er en samling af files, der på en eller anden måde udgør en enhed, fx fordi registreringssystematikken er den samme, eller fordi arkivskaberen har valgt at foretage en gruppering. Fond er en aktørs samlede arkiv.

ISAD(G) er lavet til beskrivelse af papirarkivalier, og der er på nuværende tidspunkt ikke taget officiel stilling til standardens anvendelse på elektroniske arkivalier. Men i 2000 blev der ved University College London ud fra teoretiske overvejelser og kontakter med flere dataarkiver udarbejdet en revideret version af 1994-versionen af ISAD(G), beregnet på hvad der kaldes arkiv-datasæt.⁵⁷ Genstandsområdet for beskrivelsen er således ikke arkivalier, men elektroniske datasæt, der defineres som en delmængde af elektroniske arkivalier med følgende karakteristika: ”.. collections of raw data or information which have been removed from their original computing environment (databases) and can naturally be represented as a series of tables containing columns for particular types of informations and rows for each instance of data.”⁵⁸

I det følgende vil jeg referere fra artiklen, især med vægt på ændringerne i forhold til standarden, da de afspejler de punkter, hvor e-arkivalier adskiller sig fra papirarkivalier.

Der er overordnet visse problemer med tilpasningen af de fire niveauer for arkivbeskrivelsen. Dog ikke på det øverste niveau, hvor fond også kan omfatte elektroniske data-

⁵⁷ Shepherd

⁵⁸ Shepherd, s. 57

sæt. Serie defineres som et sæt af relaterede filer, det vil sige en database. File er det største problem, både fordi det er svært at parallelisere til databaser, og fordi begrebet betyder noget andet i en IT-verden. Det frarådes derfor at bruge file, bortset fra når det skal angives, hvilket niveau i ISAD(G) en beskrivelse befinder sig på. I stedet kan file fx kaldes datasæt. Definitionen af file eller datasæt er lidt uklar, men det synes at være de enkelte afleveringer fra en database, fx en årgangsaflevering eller et øjebliksbillede. På det nederste fjerde niveau paralleliseres item med den enkelte tabel i en database.

På figur 6 ses oversigten over de elementer, der skal indgå i beskrivelsen af e-arkivalier, hvor kursivering markerer de elementer, hvor der foreslås tilføjelser eller redefinitioner i forhold til standarden (1994-versionen). Den mest gennemgående begrundelse for tilføjelserne er, at datasæt af forfatterne defineres som todimensionale, hvormed menes, at struktur, altså selve systemet eller databasen, og indhold kan skilles ad.

Denne såkaldte todimensionalitet medfører, at der i Identity Statement Area indføres flere elementer til dateringer. Der skal således være dateringer af systemets / strukturens start, af (første) datainddatering(er), af sidste datainddatering og af sidste acces. Forfatterne har en længere diskussion om, at dateringerne i praksis ofte må baseres på skøn, for hvordan defineres fx den dato, der angiver start for systemet – er det startdato for udviklingsprocessen eller startdato for systemets ibrugtagning. I nogle tilfælde vil startdatoen for data ligge før startdato for strukturen, fordi ældre data er blevet inddateret eller konverteret fra et ældre system.

Todimensionaliteten medfører også redefinitioner i Context Area, hvor elementet Names of creator(s) erstattes af et element, der kaldes statement of responsibility (3.2.3.), hvor mange forskellige aktører kan beskrives med deres roller. Det kan fx være skaberen af systemet, den ansvarlige for systemet, indsamleren af data, den ansvarlige (person eller system) for inddateringen eller brugeren af data. Også her anføres det, at det er en skønssag, hvor mange og hvor meget der skal registreres.

På samme område er elementet Administrative/biographical history suppleret med et element, der kaldes Aim and purpose (3.2.2.). Begrundelsen er, at det af hensyn til især statistisk forskning er væsentligt at få registreret så mange oplysninger som muligt om formålet med det elektroniske datasæt og om brugen af data, men at det vil blive for omfattende og uoverskueligt at registrere det under standard-element beregnet på historik.

Som endnu en følge af todimensionaliteten er indføjet Structure Area med fire elementer (3.4.). Logical structure and schema (3.4.1.) drejer sig om datasættets opbygning og filernes evt. indbyrdes relationer i en database. The active/working nature of the database (3.4.2.) er en lidt uklar benævnelse, men dækker over oplysninger om, hvorvidt data eller dele af data er afsluttede, akkumulerende eller overskrives. Hvis data er akkumulerende, bør det angives, om såkaldt historiske data, altså de data der ikke længere er gældende, gemmes i de oprindelige filer eller overføres til arkivfiler. Disse to elementer erstatter det oprindelige element System of arrangement.

- 3. ELEMENTS OF DESCRIPTION
 - 3.1 Identity Statement Area
 - 3.1.1 Reference code(s)
 - 3.1.2 Title
 - 3.1.3 Dates of creation of the material in the unit of description
 - 3.1.4 *Dates of contents of the material in the unit of description*
 - 3.1.5 *Date of last input*
 - 3.1.6 *Date of last access*
 - 3.1.7 Level of description
 - 3.1.8 Extent of the unit of description (quantity, bulk, or size)
 - 3.2 Context Area
 - 3.2.1 Administrative/biographical history
 - 3.2.2 *Aim and purpose*
 - 3.2.3 *Statement of responsibility*
 - 3.2.4 Custodial history
 - 3.2.5 Immediate source of acquisition
 - 3.3 *Scope and Content Area*
 - 3.3.1 Scope and content/abstract
 - 3.3.2 Appraisal, destruction and scheduling information
 - 3.3.3 Accruals
 - 3.4 *Structure Area*
 - 3.4.1 *Logical structure and schema*
 - 3.4.2 *The active/working nature of the database*
 - 3.4.3 *Data capture and validation before transfer to the archive*
 - 3.4.4 *Constraints on the data reliability*
 - 3.5 Conditions of Access and Use Area
 - 3.5.1 Legal status
 - 3.5.2 Access conditions
 - 3.5.3 Copyright/conditions governing reproduction
 - 3.5.4 *Data protection requirements*
 - 3.5.5 Language of material
 - 3.5.6 Finding aids
 - 3.6 Applied Materials Area
 - 3.6.1 Related units of description
 - 3.6.2 Associated material
 - 3.6.3 Publication produced by the originating body
 - 3.6.4 *Publication produced by external researchers*
 - 3.7 *Original System Attributes Area*
 - 3.7.1 *Hardware*
 - 3.7.2 *Operating system*
 - 3.7.3 *Application software*
 - 3.7.4 *User interface*
 - 3.8 *Archive Management Area*
 - 3.8.1 *Digital processing and arrangement*
 - 3.8.2 *Content validation*
 - 3.8.3 *Transformation validation*
 - 3.9 Note Area
 - 3.9.1 Note

Fig. 6⁵⁹

De to sidste elementer i struktur-området er derimod specifikke for e-arkivalier og drejer sig om datafangst, validering af data og begrænsninger i pålidelighed. Disse oplysninger karakteriseres af forfatterne som især væsentlige for statistisk brug af data.

Med henblik på at beskrive den oprindelige struktur er også indføjet en nyt område kaldet Original system attributes area (3.7.). Begrundelsen for dette er, at datasættene vil blive afleveret i et systemuafhængigt format, og at man skal tænke på, at det gerne skal være muligt at rekonstruere systemet. At det ikke er nemt at vurdere, hvad det kræver,

⁵⁹ Shepherd, s. 68

nævnes af forfatterne, og de tilråder, at så meget som muligt inkluderes. Området er delt op i elementerne hardware, operativsystem, applikationssoftware og brugergrænseflade.

Det sidste helt nye område kaldes Archive management area (3.8.), og det skal indeholde oplysninger om, hvad der er sket med data, efter de er modtaget af arkivet, da disse oplysninger er påkrævede for at understøtte autenticiteten samt give yderligere oplysninger om selve data. Området er, som det fremgår af figuren, opdelt i tre elementer, men en nærmere definition af dem er ikke givet.

Sammenfattende kan siges, at forslagene til ændringer i ISAD(G) især er begrundet med, at struktur og indhold ved e-arkivalier er adskillelige i modsætning til papirarkivalier. Det medfører, at det er de rå data, der afleveres til arkivet, mens strukturen ikke afleveres, men skal beskrives. Desuden giver de mange muligheder for statistisk brug af data andre behov for beskrivelse end papirarkivalier.

Ovenfor er nævnt, at forfatterne til tilpasningen af ISAD(G) tildels arbejdede på grundlag af kontakter med dataarkiver, hvilket først og fremmest var The National Digital Archive of Datasets (NDAD), der anvender ISAD(G), og hvis datasætbeskrivelser findes på deres hjemmeside⁶⁰. NDAD drives af University of London Computer Centre (ULCC) på vegne af det britiske nationalarkiv, og det modtager elektroniske arkivmateriale fra centrale myndigheder, på nuværende tidspunkt hovedsagelig surveys og statistikregistre.

NDAD følger ISAD(G)'s hierarkiske struktur, således at indgangen er en oversigt over afleverende myndigheder (departments) og serier, af hvilke der er registreret godt 40. Ved valg af en serie fås en oversigt over et eller flere datasæt (eller files), det vil sige de enkelte afleveringer, typisk årgangsaflæveringer. Ved valg af et datasæt fås en oversigt over tabeller i datasættet.

Til hver serie, hvert datasæt og hver tabel findes der såkaldte katalogoplysninger, der er bygget op af ISAD(G)-elementer, tilpasset e-arkivalier. Det kan umiddelbart virke lidt svært at få overblik over, på hvilket niveau hvilke oplysninger er registreret. Den tilpassede ISAD(G)-standard følges som sagt, men det er varierende, hvilke elementer der er medtaget, og hvor omfattende oplysninger der gives.

Katalogoplysningerne er fri tekst, der er struktureret ved hjælp af overskrifter, der svarer til ISAD(G)-elementerne. Det er mit indtryk, at det er oplysninger, som en eller flere medarbejdere ved arkivet har erhvervet sig gennem arbejdet med afleveringer og på denne måde videreformidler. Herudover findes der dokumentation, der stammer fra arkivskaberen, og denne dokumentation er også katalogiseret og kort beskrevet, og desuden gøres der i nogle tilfælde i den arkivskabte katalogbeskrivelse opmærksom på, at der kan hentes yderligere oplysninger i dokumentationen fra arkivskaberen.

Som nævnt er den tilpassede standard rettet mod datasæt eller databaser, der defineres som en undertype af arkivalier, mens fx dokumenter ikke er omfattet. Imidlertid er det svært at forestille sig en elektronisk dokumentsamling uden en eller anden form for tilknyttet søge-database. Således er der til NDAD afleveret et enkelt elektronisk doku-

⁶⁰ <http://www.ndad.ulcc.ac.uk>

mentarkiv inklusiv en søgebase i Access, og katalogoplysninger til dokumentarkivet er udarbejdet med de samme punkter som katalogoplysninger til datasæt.

Sammenfattende vil jeg sige, at den tilpassede ISAD(G)-standard og også eksemplerne på dens brug hos NDAD umiddelbart forekommer anvendelig som inspiration til, hvad det er, der bør dokumenteres, enten af arkivskaberen eller arkivet.

3.5. Dataarkiver

I en række lande findes dataarkiver, der typisk er opstået i tilknytning til et universitet, med den hovedopgave at bevare spørgeskemaundersøgelser for sekundær analyse især inden for samfundsvidenskabelig forskning. En del af disse dataarkiver har eksisteret i en række år og er opstået, før de nationale arkiver for alvor begyndte at beskæftige sig med e-arkivalier.

Dataarkivernes datasæt adskiller sig således fra e-arkivalier ved hovedsagelig at stamme fra spørgeskemaundersøgelser og ved at være indsamlet med henblik på sekundær analyse i den nærmeste fremtid, gerne over nettet, når de er rensat for personoplysninger. Datasættene afleveres af forskere og forskergrupper på grund af deres egen interesse i, at data bevares for sekundær analyse. En anden forskel er, at dataarkiverne hører til i en samfundsfaglig forskningstradition, mens de nationale arkiver er knyttet til en arkivistisk og (administrations)historisk forskningstradition. Det er ikke uvæsentligt, og fx medfører det forskellige begreber for noget, der måske et langt stykke hen ad vejen er det samme. Nogle dataarkiver beskæftiger sig dog også med det historiske fagområde, da de står for digitalisering eller indtastning af ældre historisk materiale.⁶¹

Der er således både reelle og historisk betingede forskelle på dataarkiver og nationalarkiver, men der er også ligheder. Nationalarkiverne modtager normalt arkivalier fra offentlige myndigheder, og når det gælder e-arkivalier, er der blandt disse også data fra spørgeskemaundersøgelser og data fra IT-systemer, hvor en del af formålet med systemet har været at lave kvantitative analyser. Jeg vil derfor kort nævne nogle af de krav, som dataarkiver stiller til dokumentation af datasæt.⁶² Mit udgangspunkt er Dansk Data Arkiv, men jeg har sammenlignet med nogle af de andre dataarkivers søgekataloger, og de indeholder stort set de samme punkter, bl.a. på grund af det internationale samarbejde der i mange år har været mellem dataarkiverne.

Dokumentationen kan deles i to. For det første studiebeskrivelsen, der er en dokumentation af undersøgelsen, der resulterede i datasættet, og for det andet variabelbeskrivelsen, der er en dokumentation af selve datasættets variable, koder og tilhørende spørgeskema.

Studiebeskrivelsen, der kan ses som en parallel til en dokumentation af oprindelse og brug, er placeret i dataarkivets søge- eller datakatalog, der har en fast struktur med overskrifter, under hvilke der skrives fri tekst. Nedenstående liste er en sammenfatning af de punkter, der indgår i søgekatalogerne, og udover disse er der punkter, som dataarkivet har tilføjet med henblik på fremfindning af datasættet, såsom referencer og søgeord.

⁶¹ Konkrete eksempler er udover DDA i Danmark beskrevet i Schürer, s. 47-48 (Tyskland), s. 55-56 (Holland), s. 65-66 (Sverige) og s. 79 (England).

⁶² Er baseret på Rasmussen, på Kjeldgaard samt oplysninger fra et besøg hos DDA.

- Involverede personer og organisationer: Donor, primær-undersøgelsesleder, dataindsamler og økonomisk sponsor.
- Emne, ofte med en ret grundig beskrivelse af undersøgelsen og datasættet.
- Datatype, fx spørgeskema, interview, registerdata.
- Observationsenheder: Type og antal
- Datasættets størrelse: Antal enheder og antal variable.
- Tidsangivelser: Afleveringstidspunkt, periode dækket af undersøgelsen, tidsrum for dataindsamling, tidsdimension.
- Definition af univers og udvælgelsesprocedurer.
- Dataindsamlingsmetode.
- Betingelser for reanalyse, herunder tekniske oplysninger og adgangsbetainger.
- Referencer til beslægtede studier.
- Referencer til publikationer fra primærundersøger.

Dokumentationen af spørgeskemaet og de koder, der anvendes i datasættet, fås i nogle tilfælde i den form, hvori det er afleveret fra primærundersøger. I andre tilfælde foretager dataarkivet en oparbejdning, således at datasættet og den tilhørende dokumentation bliver nemt tilgængelig for en bruger, samtidig med at unøjagtigheder og data, der ikke kan tolkes, opdages og søges afklaret hos primærundersøger, mens han stadig kan huske detaljer af undersøgelsen.

På baggrund af min meget overfladiske undersøgelse af dataarkiver vil jeg i forhold til dette projekts problemstilling pege på følgende:

- Dataarkivernes krav til dokumentation kan også have relevans for de e-arkivalier, der har karakter af datasæt beregnet på kvantitativ analyse. Jeg vil dog antage, at punkterne ovenfor kan indgå i en mere generel liste over, hvad der bør dokumenteres.
- Det er tankevækkende, hvor meget arbejde der i nogle tilfælde synes påkrævet for at få information ud af data, selv om der ved datasæt fra spørgeskemaundersøgelser ofte kun er en fil (tabel), samt at muligheden for at kontakte dataskaberen kan være afgørende.
- Dataarkiverne har lige fra deres tilblivelse indsamlet datasæt med henblik på, at de skal anvendes til sekundær analyse hurtigst muligt, mens nationalarkiver har haft et mere langsigtet mål og har indsamlet e-arkivalier for at gemme dem med henblik på brug engang ud i fremtiden.

4. FORSLAG TIL DEN IDEELLE DOKUMENTATION

Dette hovedafsnit drejer sig om forslag til den ideelle dokumentation af data i system-uafhængige arkiveringsversioner. Først diskuteres begrebet dokumentation, og hvem der udarbejder dokumentationen hvornår, og dernæst opstilles på skemaform den ideelle liste med samtlige dokumentationselementer.⁶³ Jeg kalder det den ideelle liste, da det næppe vil være praktisk og ressourcemæssigt realistisk at få alle elementer dokumenteret for alle arkiveringsversioner. Desuden er det en bruttoliste, hvor det ikke vil være lige relevant at få alle elementer dokumenteret for alle arkiveringsversioner.

Udgangspunktet for listen er de punkter, der blev opstillet som konklusioner i hovedafsnit 2, og nedenfor gentages disse samlet. Listen er desuden inspireret af de projekter og anbefalinger, der er gennemgået i hovedafsnit 3.

For det første blev følgende afledt af undersøgelsen af proveniensbegrebet og e-arkivalier:

- Hvem har skabt og haft ansvaret for strukturen (IT-systemet) og for overførslen til arkiveringsversion?
- Med hvilket formål er strukturen skabt?
- Hvem har haft ansvaret for indholdet?
- Hvorfor og hvordan er dataindsamling, datainddatering og evt. validering foregået?
- Hvilke data er blevet overført automatisk fra andre systemer?
- Hvem har anvendt data?
- Hvordan og i hvilken kontekst er data blevet anvendt?

For det andet blev følgende afledt af undersøgelsen af e-arkivalier set fra den infologiske model og den splittelse, der er af medie, indhold og struktur. Der skal her gøres opmærksom på, at de to første pinde drejer sig om dokumentation i forhold til bits og tegn, som ikke er en del af dette projekt. Der vil derfor på den ideelle liste ikke være nævnt dokumentation af bits og tegn – her drejer det sig om den dokumentation, der skal muliggøre omformningen af data til information.

- Bevaring af de mindste betydningsadskillende enheder, bits, ved at sikre, at lagringsmedier ikke forgår, og ved at sikre dokumentation af, med hvilket udstyr de skal læses.
- Bevaring af mulighed for omformningen af bits til de for mennesket mindste betydningsadskillende enheder, tegn, ved at sikre, at der findes programmer eller tilstrækkelig dokumentation til at kunne skabe programmer, der kan foretage omformningen.
- Bevaring af de datafragmenter (felter), der findes i det oprindelige IT-systems database, samt dokumentation af informationen i det enkelte datafragment.
- Bevaring af dokumentation af de dele af applikationen, der har skabt dataelementerne ved på brugergrænsefladeniveau at definere struktur.
- Bevaring af dokumentation af de dele af software, der har haft betydning for data i databasen. Det vil sige de beregninger, der foretages automatisk, og de dataoverførsler der sker fra andre systemer.

⁶³ Begrebet dokumentationselement stammer fra Rasmussen.

For det tredje blev følgende afledt af overvejelserne om autenticitet:

- Omformninger på bit-niveau. Når bits populært sagt flyttes fra et lagringsmedie til et andet, skal det dokumenteres, at det er kontrolleret, at der ikke er sket ændringer. Hvis bits ændres på grund af konvertering til andet tegnsæt, bør enten hver enkelt ændring eller type af ændring dokumenteres. Dette vil også gælde, hvis der sker konvertering af elektroniske dokumenter.
- Omformninger på datafragment-niveau, det vil fx sige, hvis der sker ændringer i databasestrukturen. Sådanne omformninger kan både ske, mens systemet er i drift, ved konvertering fra IT-system til arkiveringsversion og ved konvertering hos arkivet.
- Ændringer af struktur på brugergrænsefladeniveau. Hvis strukturen helt falder bort, fordi data afleveres systemuafhængigt, må strukturen dokumenteres på anden vis.

4.1. Begrebet dokumentation og formålene med dokumentationen

Indledningsvis skal der knyttes nogle bemærkninger til begrebet dokumentation, for det kan diskuteres, om det er dette begreb, der skal anvendes for 'det', der ledsager selve data, forstået som udtrækket fra en database eller de elektroniske dokumenter.

I den officielle danske arkivterminologi anvendes udtrykket 'data om arkiveringsversionen', som opdeles i generel information og metadata.⁶⁴ I OAIS anvendes begrebet information med forskellige præfikser.⁶⁵ I anbefalingerne fra ICA anvendes begrebet metadata, og hos dataarkiver anvendes begrebet (data)dokumentation. Så der er flere mulige begreber: data, information, metadata og dokumentation.

Begrebet information vil jeg ikke anvende, da det ikke passer med begrebets definition i infologien. Der er således ingen information i 'det', der ledsager dataudtrækket – det er data, der på basis af forudviden kan fortolkes til information, der kan indgå som forudviden i fortolkningen af dataudtrækket. Derfor vil det være korrekt at betegne 'det' som data, men det begreb er nok for bredt. Man kunne så vælge som i den danske terminologi at kalde 'det' data om arkiveringsversionen, men det medfører nogle overvejelser om, hvorvidt arkiveringsversionen kun er dataudtrækket, mens det ledsagende materiale, der skal muliggøre fortolkning, ligger uden for arkiveringsversionen. Metadata ville derimod være et udmærket begreb set i lyset af, at 'det' er data om selve data, men metadata er problematisk at anvende, da det i dansk arkivterminologi har fået en begrænset betydning.

Jeg har derfor valgt begrebet dokumentation, der i almindelig sprogbrug betyder "et materiale, som indsamles og fremvises med det formål at bevise noget eller danne grundlag for videre forskning." I arkivsammenhæng er det hos DDA mere præcist defineret som "dokumentation er information, data skal tilføres for at give mening."⁶⁶ Hvis definitionen skal være i overensstemmelse med den infologiske ligning, kan den lyde:

⁶⁴ Bekendtgørelse nr. 342 af 11. marts 2004 om arkiveringsversioner af bevaringsværdige data fra elektroniske arkivsystemer, bilag 3.

⁶⁵ Figur 4, s. 32.

⁶⁶ Kjeldgaard, s. 24.

Dokumentation er data, som kan fortolkes til information, der er forudviden for fortolkning af data(udtrækket).

Der er således tale om, at fortolkningen går i ring eller snarere er som en hermeneutisk cirkel, hvor det også kræver forudviden at fortolke data i dokumentationen. Spørgsmålet kan derfor være, om denne nødvendige forudviden eller dele af den også skal dokumenteres, hvilket naturligvis i praksis er umuligt. Fortolkeren har selv et ansvar for at få den nødvendige forudviden – eller han må fylde hullerne efter bedste overbevisning! Men pointen er, at det ikke er muligt at sige, at en arkiveringsversion er fuldt dokumenteret, så den er selvforklarende, for der vil altid være behov for yderligere forudviden for at fortolke de data, der udgør dokumentationen. Hos OAIIS formuleres det på denne måde:

”The recursive nature of Representation Information, which typically is composed of its own data and other Representation Information, typically leads to a network of Representation Information objects. ... The OAIIS should then make a decision between maintaining the minimum Representation Information needed for its Designated Community, or maintaining a larger amount of Representation Information that may allow understanding by a larger Consumer community with a less specialized Knowledge Base.”⁶⁷

I overskriften til dette afsnit spørges der også om, hvad formålene er med dokumentationen. Dette forekommer forhåbentlig læseren lidt tåbeligt, da det gerne skulle være fremgået af meget af det forudgående. Men jeg vil pege på, at der også kan være et andet eller sideordnet formål hos arkiver, nemlig at dokumentation kan indgå i et arkivskabt søgemiddel, fx en arkivdatabase, med det formål at gøre det nemmere for brugeren at lokalisere det e-arkivalie eller datasæt, der kan give svar på hans spørgsmål. Dette formål er især tydeligt hos ISAD(G) og dataarkiverne. Så dokumentation kan udover at give grundlag for fortolkning have et heuristisk formål, hvilket kan have betydning for udformningen og placeringen af dokumentationen.

4.2. Oprindelig eller arkivskabt dokumentation

Der er tre muligheder for, i hvilken sammenhæng og af hvem dokumentationen kan skabes:

- Dokumentation kan være oprindelig. Hermed menes, at den har haft en oprindelig værdi som dokumentation i forbindelse med udformningen og brugen af IT-systemet. En sådan dokumentation er lavet uden tanke for genanvendelsen i arkivsammenhæng, og den kan også betragtes som et arkivalie.
- Dokumentation kan være arkivskabt, hvor en arkivmedarbejder kombinerer viden om IT-systemet, dets brug og kontekst med viden om, hvad der kan være behov for at få dokumenteret af hensyn til fremfindning og fortolkning i arkivsammenhæng.

⁶⁷ Reference Model for an Open Archival Information System, s. 2-4.

- Dokumentation kan være skabt af aktøren⁶⁸ med henblik på arkiv, hvilket er det samme som punktet ovenfor, men dokumentationen udarbejdes af den aktør, der kender systemet.

Hver af de tre typer har sine fordele og ulemper, og de udelukker ikke hinanden.

De væsentligste fordele ved *den oprindelige dokumentation* er, at den findes i forvejen, og at den afspejler system, brug og organisatorisk sammenhæng på egne præmisser – som ethvert andet arkivalie. Den største ulempe er, at den ofte er meget omfattende, fx en systembeskrivelse på seks ringbind eller en arbejdsgangsbeskrivelse på to ringbind, og den bliver ikke nemmere at overskue, hvis den scannes ind. Der kan være en tilbøjelighed til at nøjes med bevaring af dele af en sådan systembeskrivelse eller arbejdsgangsbeskrivelse, vel egentlig ikke af pladshensyn, men mere fordi man mener, at den vil være uoverskuelig for en fremtidig forsker. Det vil selvfølgelig kræve en del arbejde at sætte sig ind i fx en fuld systembeskrivelse, men den bør bevares, så forskeren ikke afskæres fra muligheden, i stedet for en del-kassation på et spinkelt grundlag, fordi arkivaren heller ikke har tiden til at sætte sig ind i den.

Andre ulemper ved den oprindelige dokumentation kan være, at den er svær at lokalisere, eller at den ikke findes. Arbejdsgangsbeskrivelsen kan fx være overleveret mundtligt fra medarbejder til medarbejder som en del af kommunikationsfællesskabet.

Dele af den oprindelige dokumentation kan være publiceret. Det kan fx være love og bekendtgørelser, der fortæller om formålet med systemet, eller det kan være publikationer, der viser, hvorledes data er blevet anvendt, hvilket kan være ret omfattende ved forskningsdatabaser. I sådanne tilfælde kan overvejes henvisninger til publikationerne i stedet for omfattende indscanning. Ligeledes kan henvisning til oprindelig dokumentation i andre arkiveringsversioner overvejes, fx ved årgangsafleveringer hvor der ikke sker ændringer, og ved mange ensartede arkiveringsversioner fra enstypemyndigheder.

De væsentligste fordele ved *den arkivskabte dokumentation* er, at den viden, som arkivmedarbejderen får ved arbejdet med bevaringsvurdering og aflevering, ikke går tabt, og at denne dokumentation kan målrettes det faktum, at arkivaliet skal anvendes i en anden sammenhæng end den oprindelige. Den arkivskabte dokumentation vil ofte være mindre omfattende og nemmere tilgængelig end den oprindelige dokumentation og dermed være velegnet til at indgå i en arkivdatabase. Den arkivskabte dokumentation kan ikke erstatte den originale dokumentation, men de kan supplere hinanden. Den arkivskabte dokumentation kan dog udfylde huller, hvis oprindelig dokumentation mangler. Et problem kan være, at det kræver ressourcer hos arkivet at udarbejde dokumentationen.

Den aktørskabte dokumentation med henblik på arkiv har den fordel, at arkivet har udstukket retningslinier, der tager højde for genanvendelsesværdien, samtidig med at det er den, der kender systemet bedst, der laver dokumentationen. For det første er denne type dokumentation uundværlig, når det gælder dokumentation af arkiveringsversionens

⁶⁸ Aktøren skal her forstås bredt som en person eller organisation uden for arkivet, hvor personen eller organisationen har en viden om IT-systemet, det være sig som skaber af struktur eller indhold.

struktur og indhold. En sådan dokumentation findes ikke som oprindelig og kan vanskeligt laves af arkivet. For det andet kan en sådan dokumentation anvendes til at udfylde huller i den oprindelige dokumentation. For det tredje kan den aktørskabte dokumentation med henblik på arkiv erstatte den arkivskabte dokumentation, hvis arkivet ikke selv har ressourcer til at udarbejde den. Det må dog betragtes som en nødløsning, i hvert fald hvis dokumentationen skal indgå i en arkivdatabase, hvor det vil være hensigtsmæssigt med en vis ensartethed på tværs af arkiveringsversioner.

4.3. Den ideelle liste

På de følgende sider er opstillet den ideelle eller totale liste på skemaform i et forsøg på at operationalisere de foregående mange siders teoretiske undersøgelser og overvejelser.

Listen er inddelt i otte kolonner. I kolonne 1 er angivet fem forskellige hovedområder, som jeg har valgt at placere dokumentationselementerne indenfor: Kontekst, data / indhold, struktur, relaterede data og arkivdata. I kolonne 2 er de enkelte dokumentationselementer listet, og i kolonne 3 er de yderligere uddybet.

Det er hverken her eller i det foregående drøftet, hvilken form indholdet i dokumentationselementerne kan have, men normalt vil det være en form for tekst. Teksten kan være ustrukturerede dokumenter, hvormed menes, at der ikke er en maskinlæsbar strukturering, som muliggør søgning inden for det enkelte dokument. Sådanne dokumenter kan fx være konverteret til TIFF-dokumenter i arkiveringsversionen, hvor det heller ikke umiddelbart er muligt at foretage fritekstsøgning. Den oprindelige dokumentation vil være sådanne ustrukturerede dokumenter, hvilket sammen med deres omfang i mange tilfælde vil kræve en større arbejdsindsats af en forsker. Man kan dog forestille sig, at stadig mere oprindelig dokumentation vil have været tilgængelig via et intranet (eller internettet), og det vil muligvis være hensigtsmæssigt at bevare den opmærkning, der er foretaget i denne forbindelse. De dokumentationselementer, som skabes med henblik på arkiv enten af arkivet selv eller af aktøren, vil med fordel kunne opmærkes fx med XML, således at de enkelte elementer bliver søgbare.

Kolonne 4 –7 er afkrydsningskolonner, hvor de tre første angiver de tre muligheder for, i hvilken sammenhæng og af hvem dokumentationen er skabt, således som det er skitseret i forrige afsnit.

Der er sat kryds ved Oprindelig ved de dokumentationselementer, hvor der antages at kunne eksistere relevant oprindelig dokumentation. Der vil ikke i alle tilfælde være en skriftlig dokumentation, og der vil antageligvis også være tilfælde, hvor den må vurderes at være irrelevant for en fortolkning af data. I mange tilfælde vil den samme oprindelige dokumentation, fx en sagsbehandlerrinstruks eller en systembeskrivelse, indeholde dokumentation af flere elementer.

Der er sat kryds ved Arkivskabt ved en række dokumentationselementer. For det første ved elementer til grundlæggende navngivning og datering, som er basis for identifikation, og hvor det er hensigtsmæssigt, at arkivet udformer disse elementer på standardiseret måde. For det andet ved en række dokumentationselementer, hvor der også foreslås oprindelig dokumentation. Grunden til dette er som tidligere nævnt, at den oprindelige

dokumentation ofte vil være omfattende og tidskrævende at sætte sig ind i, mens en supplerende og kortfattet arkivskabt dokumentation på basis af den viden, som arkivmedarbejderen alligevel har erhvervet sig, vil kunne give forskeren et overblik. For det tredje ved de dokumentationselementer, der er placeret under hovedområdet Arkivoplysninger, og som har at gøre med modtagelses- og arkivkonverteringsoplysninger.

Med Arkivrettet menes den dokumentation, der er skabt af aktøren rettet mod arkivet, og der er sat kryds ved de dokumentationselementer, der drejer sig om datafragmenterne og strukturen i arkiveringsversionen. Som nævnt ovenfor kan den arkivrettede dokumentation eventuelt erstatte den originale dokumentation og den arkivskabte, men det er ikke markeret i skemaet.

Den sidste afkrydsningskolonne kaldes I arkivdatabase, hvormed menes, at den arkivskabte dokumentation kan placeres som en del af en arkivdatabase, således som dataarkiverne gør det, og således som der lægges op til i ISAD(G) og den tilhørende XML-opmærkning (EAD).

Den ottende og sidste kolonne indeholder supplerende bemærkninger til afkrydsningen.

	Dokumentationselement	Kommentar til dokumentationselement	Oprindeligt	Arkivskabt	Arkivrettet	I arkivdatabase	Kommentarer til afkrydsning
Kontekst	Navn(e) på oprindeligt IT-system			X		X	
	IT-systemets provenienser	Hvem har været ansvarlig for struktur, for dataindsamling, for indlæsning, for anvendelse. Hvis multiprovenienser skal dokumenteres, er det ikke nok med navne, også roller og sammenhænge mellem aktører og data skal dokumenteres, i nogle tilfælde helt ned på feltniveau.	X	X		X	Et valg hvilke af provenienserne, der skal registreres i arkivdatabase
	Formål med IT-systemet	Så udførligt som muligt. Vanskeligt at konkretisere yderligere, da det bl.a. afhænger af, om IT-systemet anvendes i en administrativ eller forskningsmæssig sammenhæng. Også eventuelle ændringer over tid	X	X		X	
	Proces-kontekst	Hvordan IT-systemet er indgået i den organisatoriske kontekst, i arbejdsprocesser og forretningsgange. Også eventuelle ændringer over tid	X	X		X	Arkivskabte skal være meget overordnet og kan være vanskeligt at skille fra formål

	Dokumentationselement	Kommentar til dokumentationselement	Op-rin-delig	Ar-kivs-kabt	Ar-kiv-ret-tet	I ar-kiv-data base	Kommentarer til afkrydsning
	Population	Beskrivelse af hvilke typer enheder (fx individer), der er det centrale i systemet, og hvad der karakteriserer dem	X	X		X	Især ved databaser til forskningsformål, hvor der vil være oprindeligt materiale, eller aktør kan skabe det. Vanskeligere ved adm. systemer
Data / indhold	Første og sidste dato for ind-datering / indlæsning			X		X	
	Første og sidste dato for data	Kan variere fra inddatering / indlæsning, fx hvis ældre data indtastes / indscannes.		X		X	
	Dataindsamling	Hvem, hvordan og hvornår (overlapning til proveniens)	X	X		X	
	Datainddatering og indlæsning	Hvem og hvordan (overlapning til proveniens). Også automatisk overførsel fra andre systemer. Kan i nogle tilfælde være relevant at angive på felt-niveau	X	X		X	
	Pålidelighed af data	I hvilket omfang er der foretaget kontrol af pålideligheden	X	X		X	
	Brug af data	Af hvem, til hvad og hvordan (overlapning til proveniens)	X	X		X	

	Dokumentationselement	Kommentar til dokumentationselement	Op- rin- delig	Ar- kivs- kabt	Ar- kiv- ret- tet	I ar- kiv- data base	Kommentarer til afkrydsning
	Forklaringer til det enkelte fragment (felt)	Således at det kan fortolkes, hvilken fragmentarisk information det enkelte datafelt indeholder	X		X		Det oprindelige fx datadictionary. Meget væsentligt med forklaringer til felter i arkiveringsversionen
	Kodeforklaringer	En del af forklaringen til det enkelte datafragment, men udskilt som selvstændigt punkt, da det er absolut påkrævet	X		X		
	Beregningsforklaringer	Forklaring til de datafragmenter (felter), der er fremkommet ved beregning	X		X		Vil ofte være en del af funktionaliteten. Derfor væsentligt med forklaringer til felter i arkiveringsversionen
	Konverteringsoplysninger	Konverteringer, mens data er i brug og ved overførsel til arkiveringsversion – først og fremmest hvis det har medført ændringer af data	X		X		Evt. oprindelig ved konverteringer i brug. Væsentligst er arkivrettet ved konvertering til arkiveringsversion
Struktur	Databasestruktur	Også ændringer over tid	X		X		

	Dokumentationselement	Kommentar til dokumentationselement	Op- rin- delig	Ar- kivs kabt	Ar- kiv- ret- tet	I ar- kiv- data base	Kommentarer til afkrydsning
	Brugergrænseflade - inddatering	Selve billedet er ikke det vigtigste, men koblingen af datafragmenter til dataelementer. Skærbilleder kan være udgangspunktet, men væsentligt, at det fremgår fra hvilke felter i databasen data stammer, og – endnu væsentligere – at det fremgår, fra hvilke felter i arkiveringsversionen	X		X		Fx brugervejledninger og systembeskrivelser, men ikke tilstrækkelig uden kobling til arkiveringsversion
	Brugergrænseflade - uddata	(Kommentar som ovenfor)	X		X		
	Brugergrænseflade - udskrifter	(Kommentar som ovenfor) Kan omfatte standardbreve, hvori data indflettes.	X		X		
	Konverteringsoplysninger	Mens systemet er i brug, og især ved overførsel til arkiveringsversion, hvis der sker ændringer	X		X		
	Oprindelig hardware og software	Ikke nødvendigt for fortolkning af data	X				
Relaterede materialer	Publikationer	Publikationer hvori data er blevet anvendt		X			
	E-arkivalier eller papirarkivalier	Henvisning til dokumentationselementer i andre arkivalier. Henvisning til arkivalier med beslægtede data (fx tidsserier eller søgemiddelrelation)		X		X	

	Dokumentationselement	Kommentar til dokumentationselement	Op-rin-delig	Ar-kivs-kabt	Ar-kiv-ret-tet	I ar-kiv-data base	Kommentarer til afkrydsning
Arkivoplysninger	Bevarings- og kassationsoplysninger	Bevaring/kassation af dele af data i e-arkivaliet		X			
	Modtagelsesoplysninger	Oplysninger fra modtagelsen og kontrollen af e-arkivaliet. Bør i nogle tilfælde indføres i oplysninger om konvertering til arkiveringsversion		X			
	Adgangsbetingelser	Ikke nødvendigt for fortolkning af data, men medtaget for fuldstændighedens skyld, jf. ISAD(G) og dataarkiver		X		X	
	Arkivkonverteringer - medie	Hvornår og hvordan. Eventuelle ændringer		X			
	Arkivkonverteringer – data (tegnset)	Hvornår og hvordan. Eventuelle ændringer		X			
	Arkivkonverteringer - struktur	Hvornår og hvordan. Eventuelle ændringer		X			

5. FORSLAG TIL RETNINGSLINIER FOR DOKUMENTATION I DET PRAKTISKE ARBEJDE MED E-ARKIVALIER

I dette hovedafsnit vil jeg komme med forslag til mere præcise retningslinier end de nugældende diffuse i afleveringsbekendtgørelsen for, hvilken dokumentation der skal følge data- og dokumentudtræk i arkiveringsversioner. Forslagene er baseret på den ideelle liste, som diskuteres i forhold til de aktuelle betingelser for arbejdet med indsamling af e-arkivalier. Forslagene til retningslinier er i første omgang tænkt som et diskussionsoplæg, men det kan selvfølgelig overvejes, om der skal foretages ændringer i bekendtgørelsen. Det vil jeg dog ikke tage stilling til her.

Dog vil jeg foreslå, at de nuværende udtryk 'data om arkiveringsversionen' og især 'generel information' ændres til dokumentation, opdelt i fx dokumentation af formål og oprindelse, af data og af struktur.

Det overordnede formål med dokumentationen set ud fra den infologiske forståelsesramme er at sikre, at det vil være muligt at fortolke data også i en fjern fremtid. Men som det fremgår af den ideelle listes kolonne "I arkivdatabase" medtænkes som formål, at det skal være muligt for en forsker at vurdere, om det netop er dette e-arkivalie, der sandsynligvis kan besvare hans spørgsmål. Dette vil jeg kalde heuristisk tilgængeliggørelse, hvormed menes, at data ikke bliver tilgængelige af det, men arkiveringsversionen bliver tilgængelig i den traditionelle papirarkivalie-terminologi, hvor tilgængeliggørelse for mig at se betyder, at arkivaliet kan findes. At dette formål medtænkes, skyldes inspiration fra dataarkiver og ISAD(G).

5.1. Oprindelig dokumentation (aktørskab)

Den oprindelige dokumentation er den dokumentation eller de arkivalier med relation til IT-systemet, der allerede findes. Praksis i dag er, at arkivaren spørger myndigheden eller aktøren, hvad der findes af materiale om formål, indhold og brug, hvorefter arkivaren efter bedste evne afgør, hvad der skal indgå i arkiveringsversionen. Der er imidlertid ikke på nuværende tidspunkt konkrete retningslinier for udvælgelsen, og den ideelle liste foreslås derfor anvendt som inspiration eller checkliste. Selve praksis med myndighedens fremfindning og arkivarens BK-vurdering af oprindelig dokumentation kan og skal der grundlæggende ikke ændres på. Baseret på egne erfaringer er der dog for mig at se nogle problemer:

- Det kan være svært på stående fod ved et afleveringsmøde at få præsenteret al dokumentation.
- Der er meget lidt tid til at forholde sig til dokumentationen og til at vurdere, om der er væsentlige huller i den, som skal udfyldes med arkivrettet eller arkivskabt dokumentation.⁶⁹
- Der er en tendens til at kassere i dokumentationen, for at den ikke skal blive for uoverskuelig (mange TIFF-sider).
- Det kan være svært at vurdere, hvor grænsen er for, hvilken dokumentation der skal indgå i arkiveringsversionen.

⁶⁹ I årsplan for både 2004 og 2005 er der to normer på henholdsvis 10 og 20 timer til alt arbejde med afleveringsforberedelse af et register.

Med hensyn til de to første problemer kan der i startbrev og ved aftale om møde lægges op til, at dokumentationen fremfindes og sendes på forhånd, altså vægte resten af dokumentationen på linje med dokumentationen af databasens struktur. I nogle tilfælde vil det næppe lykkes, men så må der være så meget mere tid til afleveringssagen, at der kan holdes et ekstra møde, efter at jorden så at sige ved det første møde er gødet for en forståelse af, hvad det drejer sig om.

Med hensyn til tendensen til kassation for at undgå for omfattende dokumentation er det rigtigt, at f.eks. en systembeskrivelse på seks ringbind både er besværlig at overskue i TIFF-format og besværlig at indscanne. Men jeg finder det problematisk at foretage delkassation uden et grundigt kendskab til både data og dokumentation. Samtidig ved vi ikke, hvordan e-arkivalierne vil blive anvendt i en nærmere eller fjernere fremtid, og hvilken dokumentation en fremtidig forsker vil have behov for. Så der bør ikke kasseres i den oprindelige dokumentation. For den forsker, som ikke vil bruge uger på at sætte sig ind i dokumentationen for at lave et enkelt opslag, kan den arkivskabte og den arkivrettede dokumentation forhåbentlig være en hjælp.

Det kan dog overvejes at undlade indscanning af materiale, der er publiceret med ISBN-nummer og ligeledes love, bekendtgørelser og cirkulærer. I stedet kan der som en del af den arkivrettede dokumentation laves en liste med henvisninger.

Hvor grænsen går for, hvilken dokumentation der skal indgå i arkiveringsversionen, kan kun være et skønsspørgsmål. Men det må antages som tommelfingerregel, at jo mere specifikt dokumentationen er knyttet til IT-systemets data, struktur og brug og til kommunikationsfællesskabet, jo mindre sandsynligt er det, at der bevares tilsvarende oplysninger andetsteds. Det gælder især dokumentationselementerne under data / indhold og struktur.

Det betyder imidlertid ikke, at dokumentationselementerne under kontekst ikke er væsentlige, men på dette område kan det være meget vanskeligt at trække grænsen. Der kan fx findes mere eller mindre omfattende journalsager omkring oprettelse, udvikling og brug af et register - et eksempel sat på spidsen kan være CPR-kontorets journalsager, der for en stor dels vedkommende er dokumentation af CPR-registret. Her er formodentlig én grænse, for sådanne sager vil normalt ikke blive inkluderet i dokumentationen. Problemet er imidlertid, at arkivaren ikke ved, om sagerne vil blive bevaret, når papirarkivet på et senere tidspunkt bliver bevarings- og kassationsbehandlet. Jeg har ikke haft mulighed for at undersøge det nærmere, men spørgsmålet er, om denne type sager betragtes som intern administration og drift og derfor kasseres.⁷⁰ Det bør derfor være muligt i forbindelse med aflevering af e-arkivalier at sikre, at papirsager, der er nødvendige som dokumentation, bevares.

5.2. Arkivrettet dokumentation (aktørskabt)

Den arkivrettede dokumentation er den dokumentation, der er skabt af en aktør uden for arkivet efter arkivets anvisninger. Som det fremgår af den ideelle liste, drejer det sig for det første om forklaringer til datafragmenterne i arkiveringsversionen. For det andet om

⁷⁰ Da CPR-kontoret er brugt som eksempel, skal det nævnes, at en stor del af dette arkiv er bevaret (indtil videre frem til 1982).

koblingen af datafragmenter til dataelementer ved dokumentation af brugergrænseflader relateret til arkiveringsversionen. Og for det tredje om databasestruktur og konvertering til arkiveringsversion. Hertil kommer, at arkivrettet dokumentation kan erstatte ikke-eksisterende oprindelig dokumentation.

For nogle af elementer er det allerede et krav eller praksis, at der skal udarbejdes arkivrettet dokumentation. Forklaringer til det enkelte datafragment indgår således i metadata som feltinfo, og koder skal forklares i tabeller i dataudtrækket, i metadata eller i den såkaldte generelle information. Der er derimod ingen krav om, at felter, hvor data er fremkommet ved beregning, skal dokumenteres. Det vil kunne beskrives i feltinfo og er muligvis også blevet det i nogle tilfælde, men jeg har ikke set eksempler på det. Det vil formodentlig være meget forskelligt, hvor stort behovet er for dokumentation af beregninger, afhængig af hvad systemet har været brugt til. Men jeg kan undre mig over, at der efter hvad jeg ved ikke har været overvejelser om problematikken i betragtning af, at der har været megen fokus på nødvendigheden af dokumentation af formler i regneark i ESDH-systemer.

Også når det gælder dokumentation af konvertering til arkiveringsversion, laves en sådan allerede, i hvert fald for arkiveringsversioner der følger standardafleveringsaftaler for journaler og ESDH-systemer. Som tidligere nævnt skal dokumentation af konverteringer ikke primært ske af hensyn til fortolkning af data, men af hensyn til autenticitet, så omfanget af den nødvendige dokumentation afhænger meget af, hvor store ændringer der sker i data og struktur. I nogle tilfælde kan det være tilstrækkeligt med en kortfattet dokumentation af, at der ikke er sket ændringer. I andre tilfælde kan der være behov for dokumentation af, at der er ændret på data i den betydning, at fx datatypen er en anden end i det oprindelige system, eller at datoer er ændret fra et andet format. Når det gælder databasestrukturen kan der være behov for dokumentation af, at der er skabt relationer, som ikke findes i det oprindelige system, eller at relationer er udeladt, fordi de ikke lever op til kravene til arkiveringsversioner.

Når det gælder dokumentation af sammenkoblingen af arkiveringsversionens datafragmenter til dataelementer til erstatning for IT-systemets skærbilleder og udskrifter, dokumenteres det kun i meget begrænset omfang som enkelte SQL-forespørgsler i arkiveringsversioner af journaler og ESDH-systemer.

Der er imidlertid et stort behov for en meget bedre dokumentation af koblingen af datafragmenter til dataelementer. Begrundelsen er for det første teoretisk og baseret på den foregående infologiske analyse, hvor det påpeges, at det først er muligt at fortolke data til information, når datafragmenterne er koblet sammen til dataelementer, hvilket i et IT-system sker på skærmen og i udskrifter – det vil sige som en del af den funktionalitet, der ikke bevares ved en systemuafhængig arkiveringsversion.

Begrundelsen er for det andet praktisk og baseret på nogle få erfaringer uden for dette projekts regi, hvor der var et konkret formål med at sammenkoble data. Det drejer sig om et system af en type, der tidligere er blevet benævnt et register med sager, hvilket vil sige en database, der bliver brugt til styring af sagsbehandling og ressourcer, men også som søgemiddel til sager, uden at det er et traditionelt journalsystem med sager og akter. Arkiveringsversionen af det pågældende system indeholder mange data, fordelt på

godt 50 tabeller, og der vil antagelig kunne findes en del oplysninger om fx den pågældende myndigheds sagsgange. I det konkrete tilfælde skulle det kun anvendes som søgemiddel, og jeg havde fået oplyst, hvilke oplysninger der skal kunne søges på, og at søgeresultatet selvfølgelig skal være et sagsnummer.

Det var temmelig tidskrævende at identificere de enkelte datafragmenter, der skulle indgå i forespørgslerne, da der som sagt er godt 50 tabeller med mellem 5 og 50 felter, men tiden er, hvad det er – arkivalier siges at skulle tage tid. Det altoverskyggende problem er, at der ikke er noget grundlag i arkiveringsversionen for en bekræftelse af, at det er de rigtige datafragmenter, jeg har koblet sammen – det er gætværk. I den medfølgende oprindelige dokumentation er der ganske vist skærbilleder, som viser nogle søgninger, der tildels svarer til de søgninger, jeg har forsøgt at lave. Men det er ikke muligt at overføre dem til arkiveringsversionen, fordi arkiveringsversionens database anvender andre tabel- og feltnavne end vist på skærbillederne, bl.a. på grund af de krav der stilles til navngivning i arkiveringsversioner.

I dette konkrete tilfælde anser jeg det dog for næsten forsvarligt at anvende forespørgslerne. For det første er et par af dem kontrolleret i det oprindelige system, der stadig er i drift. For det andet er det generelt ikke problematisk at kontrollere korrektheden af søgninger i elektroniske søgemidler, det vil sige journalsystemer, ESDH-systemer og – som i dette tilfælde – registre med sager. De informationsobjekter, man ønsker information om, er sager, og de findes stadig, så man kan bogstavelig talt gå ud til hylden og kontrollere, at søgningen var korrekt.

Det vil man imidlertid ikke kunne gøre, når det gælder registre, hvor informationsobjekterne er døde og borte om nogle år, og hvor de kort tid efter dataregistreringen kan have ændret tilstand.

Det kan derfor forekomme paradoksalt, at det ved arkiveringsversioner af elektroniske journaler og ESDH-systemer nu kræves, at der afleveres nogle få SQL-sætninger, der skal illustrere myndighedens brug af systemet, mens der ikke er et tilsvarende krav ved arkiveringsversioner af registre og databaser.

Hvis en fremtidig forsker fx vil anvende CPR / DNK, vil han få omkring 70 tabeller, der er en skønsom blanding af de centrale personregistreringstabeller, historiktabeller, DNK-tabeller, kodetabeller og tabeller med oplysninger om forskellige distriktinddelinger af landet. Hvis det er den centrale personregistrering, han er interesseret i, vil der på grund af bl.a. historik være flere tabeller med personoplysninger, og han må ud fra datafragmentdokumentation, det vil sige tabel- og feltinfo, gætte sig frem til, i hvilke tabeller og felter de data, han har brug for, er registreret. Hertil kan siges, at en vis usikkerhed ved anvendelse af kilder er et grundvilkår, og usikkerheden er her blot blevet en del større, hvilket forskeren må tage højde for, måske udvikle en ny form for kildekritik. Omvendt må det siges, at når der anvendes mange ressourcer på at sikre bevaring af bits og datafragmenter, er det uhensigtsmæssigt, at værdien forringes – hvor meget er svært at vurdere, så længe e-arkivalierne ikke bruges - hvis dataelementer ikke søges bevaret.

Hertil kommer spørgsmålet om, hvorvidt arkivet vil kunne garantere andet end autenticiteten af bits og datafragmenter? Vil arkivet kunne lave forespørgsler og søgninger, der

stilles til rådighed for forskere eller anvendes af arkivet selv til dokumentationsvirksomhed? Når det gælder søgemidler, mener jeg, at det muligt, fordi forespørgslerne som nævnt ovenfor kan verificeres ved at kontrollere på hylden, at det er den rigtige sag, der er fundet. Når det gælder registre og databaser, vil det ikke være muligt eller i hvert fald være nødvendigt med forbehold for autenticiteten, hvis der ikke i arkiveringsversionen er dokumentation, der kan anvendes som bekræftelse af, at det er de rigtige datafragmenter, der sættes sammen.

Den dokumentation der for mig at se er grundlæggende for at kunne få information ud af data i en arkiveringsversion, men som samtidig er den dokumentation, der vil være forsvundet, hvis den ikke sikres sammen med arkiveringsversionen, er således de tre elementer på den ideelle liste, der drejer sig om brugergrænseflade. Spørgsmålet er, hvordan det kan gøres, og jeg kan pege på følgende muligheder:

- At der afleveres oprindelig dokumentation i form af skærbilleder fra originalsystemet, hvortil føjes referencer (tabelnavn og feltnavn) til arkiveringsversionen. Standardudskrifter dokumenteres med standardteksterne, fx flettedokumenter, samt referencer til de tabeller og felter i arkiveringsversionen, hvorfra der hentes data til indsættelse.
- At der afleveres standardsøgninger, fx som SQL-lignende beskrivelser. Det er også her væsentligt, at der er reference til arkiveringsversionen. Skærbilledet vil mangle, hvilket kan være uheldigt, hvis væsentlige datafragmenter er en del af skærbilledet. I andre tilfælde vil der ikke have eksisteret egentlige skærbilleder, fx ved forskningsdatabaser og statistiske databaser. Vil ikke være tilstrækkeligt ved standardudskrifter.
- At der laves egentlig opmærkning af forespørgsler, som der nu gøres for journaler. Det vil måske gøre det muligt med automatisk omformning. Samme problemer som ovenfor, hvis datafragmenter, der er væsentlige for fortolkning, findes på skærbilleder eller i udskrifter.

5.3. Arkivskabt dokumentation

I dag udarbejdes der ikke arkivskabt dokumentation. Der udarbejdes dog altid en afleveringsbestemmelse, der er rettet til aktøren, ligesom der ofte i afleveringssager vil være oplysninger, der har karakter af dokumentation. Men den vil være tilfældig, og den er sammenvævet med sagsbehandlingsoplysninger af midlertidig værdi. Hvorfor det er sådan, kan man kun gisne om, men der synes implicit at være en anden indgangsvinkel end den, der ligger i ISAD(G)'s og OAIS's anbefalingerne og dataarkivernes arbejde, hvor brugen og tilgængeliggørelsen medtænkes fra starten.

Formålet med arkivskabt dokumentation er som tidligere nævnt at få en dokumentation, der er mere overordnet og nemmere tilgængelig end den oprindelige dokumentation og velegnet til at indgå i arkiv søgemidler. Samtidig gør den det muligt også på meget langt sigt at bevare den forudviden, som arkivaren erhverver sig gennem sit arbejde.

I den ideelle liste er der to typer af dokumentationselementer, hvor der er sat kryds ved arkivskabt. For det første er der en række elementer, hvor der også foreslås bevaret oprindelig dokumentation, men hvor den arkivskabte foreslås placeret i en arkivdatabase som en hjælp til identifikation og en hjælp til den forsker, der fx for at foretage nogle enkelte opslag ikke har behov eller mulighed for at bruge flere uger på at sætte sig ind i en omfattende oprindelig dokumentation. For det andet er der sat kryds ved elementer med dokumentation af, hvilke ændringer arkiveringsversionen har undergået efter modtagelse. Denne dokumentation skal ikke placeres i arkivdatabase, men i tilknytning til arkiveringsversionen, da dens hovedformål er at bekræfte autenticitet.

Den første type dokumentation kan laves i forbindelse med afleveringen, men kan også tildels udarbejdes i forbindelse med bevarings- og kassationsbehandlingen, således at den viden, der indsamles i den forbindelse med det samme bliver dokumenteret. Den anden type dokumentation skal laves i forbindelse med test og ved de konverteringer, som arkivet over tid foretager.

5.4. Placering af dokumentationen

Den oprindelige og den arkivrettede dokumentation kan placeres som hidtil og beskrives i den såkaldte geninfo.tab. Når det gælder den oprindelige dokumentation, kan den næppe navngives og beskrives mere stringent end i dag. Når det gælder den arkivrettede dokumentation, kan der overvejes en stringent navngivning ud fra de foreslåede elementer, men den gevinst der kunne være ved ensartede oversigter er nok for lille i forhold til det besvær, det vil give at få den implementeret.

Det arkivskabte - bortset fra dokumentation af test og af konverteringer - foreslås placeret i eller i tilknytning til en arkivdatabase. Med vores kendskab til DAISY må vi dog nok se i øjnene, at det ikke er realistisk i en overskuelig fremtid. Men en vis struktur og søgbarhed vil være ønskelig.

Jeg vil derfor foreslå et XML-opmærket dokument med opmærkning af de enkelte elementer jævnfør den ideelle liste. Der skal ikke nødvendigvis være lange tekster til hvert eneste af elementerne, mellem et par linier og en halv side. Dette dokument oprettes allerede ved BK-vurderingen og placeres et eller andet sted på fx s-drevet. I forbindelse med afleveringssagen findes det frem og anvendes som en hjælp til at sætte sig ind i sagen, og i løbet af afleveringssagen revideres og udbygges det. Når arkiveringsversionen er endelig godkendt, lagres dokumentet som et dokument i Generel information med fx titlen Arkivskabt. XML-opmærkningen vil forhåbentlig sikre, at dokumentet i en nærmere eller fjernere fremtid kan flyttes over i DAISY eller være tilgængelig og søgbar fra DAISY.

Ligeledes oprettes et dokument, måske også XML-opmærket, til placering af oplysninger fra testen, som anses af betydning for den fremtidige brug af arkiveringsversionen. Hvis myndigheden for at få godkendt arkiveringsversionen ændrer i forhold til det oprindelige system, bør dette dog dokumenteres i den arkivrettede konverteringsbeskrivelse. Det er jo tilfældigt, om ændringer sker, fordi myndigheden er god til at forstå bestemmelserne, og arkivaren har givet en grundig vejledning, eller om problemer først opdages ved test.

Når der senere sker konverteringer, oprettes der arkivskabte dokumenter, der dokumenterer dette. Eventuelt kan det være i form af henvisninger til en dokumentation uden for den enkelte arkiveringsversion, hvis der er tale om en samlet konvertering af mange arkiveringsversioner.

5.5. Ressourcer

De væsentligste forslag i det foregående kan sammenfattes til følgende punkter, hvori ligger en implicit rangordning:

- Sikre bevaring af alle eller et udvalg af skærbilleder, forespørgsler og udskrifter, med tabel- og feltnavne, der henviser til arkiveringsversionen.
- For hver arkiveringsversion lave et struktureret, arkivskabt dokument til fastholdelse af viden, indsamlet i forbindelse med BK-vurdering og aflevering, samt et tilsvarende dokument til fastholdelse af oplysninger fra testen.
- Anvende en checkliste som den ideelle liste som vejledning til afgørelse af, hvilken dokumentation der skal indgå i en arkiveringsversion.
- Hellere bevare for meget end for lidt oprindelig dokumentation samt sikre, at fx journalsager med oprindelig dokumentation bevares.
- Som hidtil være opmærksom på dokumentationen af datafragmenterne, evt. også når det gælder beregninger.

Det første punkt er det absolut mest ressourcekrævende for såvel aktører som arkiv. Hvor ressourcekrævende kan ikke vurderes uden praktiske erfaringer, ligesom det vil afhænge af, om alt skal dokumenteres, eller om man nøjes med det mest nødvendige, hvordan end det så skal defineres.

Der vil også være stor forskel på systemer. Nogle vil være forholdsvis nemme, mens andre vil være så omfattende, at opgaven kan synes uoverskuelig – uheldigvis er det nok også dem, hvor det vil være meget vanskeligt at uddrage information uden dokumentationen. Måske kan der til gengæld spares på ressourcer hos både aktører, testværksted og arkivarer på udformningen af relationer, hvis disse fremgår af en SQL-opmærkning. Men det er den del af dokumentationen, der vil kræve flest overvejelser og forsøg, og jo før man kommer i gang des bedre. Da det vil kræve ressourcer af myndighederne, vil det ligeledes være hensigtsmæssigt med en hjemmel i bekendtgørelsen.

De øvrige punkter kan nemmere implementeres i sagsbehandlingen. Den større opmærksomhed på den oprindelige dokumentation vil kræve lidt flere ressourcer af såvel aktører som arkivarer, og udarbejdelsen af arkivskabt dokumentation vil kræve tid både ved BK-behandlingen og afleveringssagen.

6. FORSLAG TIL DIGITALISERING AF DOKUMENTATION AF ÆLDRE E-ARKIVALIER

I dette hovedafsnit vil jeg komme med forslag til, hvorledes dokumentationen, der tidligere blev afleveret på papir, kan digitaliseres. Forslagene skal ses i sammenhæng med forslagene til dokumentation i hovedafsnit 4 og 5. Som nævnt i indledningen har jeg i projektet valgt et teoretisk udgangspunkt i stedet for at undersøge konkrete ældre e-arkivalier. Nedenstående skal derfor ses som generaliseringer, der ikke er underbygget af empiriske undersøgelser i dette projekts regi.

Indtil år 2000 blev kun selve dataudtrækket afleveret i elektronisk form, mens den medfølgende dokumentation blev afleveret på papir. Efter afleveringen blev der af arkivet udarbejdet en dokumentationsoversigt, der både findes i tilknytning til dokumentationen og i ringbind med samtlige dokumentationsoversigter. Der findes cirka 700 afleveringer med papirdokumentation, hvoraf den ene halvdel er afleveret fra 1998 og fremad, mens den anden halvdel er afleveret før 1998, de tidligste tilbage i 1970'erne.

Der er lavet et skel ved 1998, fordi der på dette tidspunkt blev udfærdiget et regelsæt, der fastsatte krav til lagringsmedie, tegnsæt og datatyper samt til dokumentationen. Den skulle for det første bestå af en filbeskrivelse med oplysning om de enkelte filers navn og indhold / funktion samt oplysninger om felternes længde, type og indhold. Herudover skulle det angives, hvis et felt var primær eller sekundær nøgle. Det skulle også angives, hvis et felt var kodet, og der skulle vedlægges kodefortegnelser. Ligeledes blev det i regelsættet i generelle vendinger beskrevet, hvilken øvrig dokumentation, kaldet opgavedokumentation, der skulle følge data, bl.a. dokumentation af det oprindelige system og administrative beskrivelser.

Indholdsmæssigt er kravene til såvel fil- som opgavedokumentation i al væsentligt de samme i 1998-reglerne som i regelsættet fra 2000, ligesom de tilladte datatyper efter 2000-reglerne er de samme som i 1998-reglerne (hvor de kaldtes værdirepræsentation). Forskellen på regelsættene er kravene til formen, hvor opgavedokumentationen efter 2000-regler skal findes som elektroniske dokumenter (kaldet generel information), og fildokumentationen (fra 2000 kaldet metadata) skal være maskinlæsbar med en XML-lignende opmærkning.

På grund af de mange ligheder mellem 1998-regler og 2000-regler kan der ved forslag til digitalisering skelnes mellem afleveringer i henhold til 1998-regler og tidligere afleveringer. Nogle af afleveringerne i årene op til 1998 vil muligvis også de facto i vidt omfang følge 1998-reglerne.

Forslagene til digitalisering afhænger også af, hvad formålet er med den, hvor der for mig at se kan være to ret forskellige. Det ene formål er, at man vil undgå, at dokumentationen skal hentes ude på arkivets hylder og i stedet umiddelbart skal kunne tilgås fra computeren. Det er den løsning, der er valgt ved de få e-arkivalier, der på nuværende tidspunkt er umiddelbart tilgængelige på en PC på Rigsarkivets læsesal. Det andet formål er, at arkiveringsversioner efter 1998-regler og ældre skal oparbejdes, så de følger regelsættet fra 2000 eller rettere det nyeste regelsæt fra 2004 – i praksis betyder det dog næsten intet, om det er regelsættet fra 2000 eller 2004, der følges, da de med hensyn til

udformningen af selve arkiveringsversionen er stort set ens. Grunden til at foretage oparbejdning er, at alle arkiveringsversioner så vil være fuldt elektroniske og ensartede, hvilket må antages at gøre tilgængeliggørelse og bevaring lettere.

Hvis formålet med digitalisering kun er at gøre det nemmere at få fat på dokumentationen, er løsningen forholdsvis enkel, om end den vil kræve noget tid. Papirdokumentationen skal for det første naturligvis scannes ind, det vil sige, at samtlige sider, der hører til en aflevering, skal køres gennem en scanner. Det er dog ikke altid helt enkelt at vurdere, hvilken dokumentation der hører til en aflevering, fordi der ved gentagne afleveringer fra samme system ofte kun er afleveret supplementer til dokumentationen af tidligere afleveringer, så spørgsmålet er, om den tidligere dokumentation også skal indgå. For det andet skal de indscannede dokumenter placeres i et antal biblioteker, der skal tildeles meningsgivende navne.⁷¹ Udgangspunktet for inddelingen og navngivningen kan være den eksisterende dokumentationsoversigt, men det kan også være en ide at foretage en yderligere opdeling for at gøre dokumentationen mere overskuelig. Hvis det er for vanskeligt at navngive bibliotekerne på meningsgivende måde, kan dokumentationsoversigten udbygges med oplysninger om, i hvilke biblioteker de enkelte dele findes.

Hvis formålet med digitaliseringen er at oparbejde arkiveringsversionerne til 2004-regler, skal der laves et større arbejde. Det kan inddeles i for det første det nødvendige og for det andet det ønskelige på basis af de foregående afsnits forslag til dokumentation.

Den eksisterende papirdokumentation skal indscannes og beskrives tilsvarende løsningen ovenfor. Forskellen er, at navngivningen af de biblioteker, hvori dokumentationen placeres, er mindre væsentlig, da indholdet af dem skal beskrives i den såkaldte geninfo.tab, hvor der er 256 tegn til rådighed. Det bør her også angives, om den indscannede dokumentation er oprindelig, arkivrettet (fx filbeskrivelser) eller arkivskabt (fx oplysninger fra testen).⁷²

Herudover skal der udarbejdes metadata, hvilket ikke kan gøres maskinelt, men vil kræve indtastning fx ved hjælp af TEA.

Når det gælder afleveringer i følge 1998-regler, bør tabel- og feltnavne, type og længde samt oplysninger til tabel- og feltinfo fremgå af filbeskrivelsen, og disse oplysninger antages uden de store problemer at kunne indtastes som metadata. Koder vil kunne være arbejdskrævende, hvis der er mange af dem. De vil være definerede i filbeskrivelsen eller i kodefortegnelser, men det skal besluttes, om de skal indtastes i metadata eller fortsat kun findes i de oprindelige kodefortegnelser. Ligeledes skal det besluttes, hvad der skal ske med definitioner af primær- og sekundærnøgler, som ikke lever op til de nugældende kriterier. Skal det accepteres, at den oparbejdede arkiveringsversions relationer ikke kan godkendes i henhold til en TEA-test, eller skal man forsøge at tilpasse

⁷¹ Den dokumentation, der er scannet ind i tilknytning til de umiddelbart tilgængelige e-arkivalier på PC'en på Rigsarkivets læsesal, er navngivet med en ikke-betydningsbærende nummerering. Det gør det meget vanskeligt at finde rundt i en dokumentation, der i papirversionen fylder flere arkivæsker.

⁷² Oplysninger fra testen er tidligere blevet placeret sammen med dokumentationen, men det har ikke været muligt siden 2000-regler. Dokumentationen i tilknytning til testen er dog også blevet lavet efter 2000, men findes nu i stedet i sagen.

relationerne. Umiddelbart vil jeg mene, at det er betænkeligt at ændre på de relationer, der er defineret af den oprindelige aktør, og hvis det gøres, skal det dokumenteres.

Når det gælder afleveringer før 1998-regler, bliver oparbejdelsen af metadata mere kompliceret, da datatyper ikke vil være i henhold til gældende regler, og da strukturen ikke vil være relationel, men hierarkisk, og oparbejdelsen må derfor ses i sammenhæng med format- og strukturkonvertering.⁷³ De metadata, der udarbejdes, skal naturligvis beskrive data med de datatyper og den struktur, som er gældende efter konverteringen. Den oprindelige fildokumentation vil kunne ses i den indscannede dokumentation, men vil ikke kunne bruges i forhold til de konverterede data, men anvendes sammen med dokumentationen af konverteringen, hvis man ønsker at undersøge autenticiteten.

Ovenstående er den nødvendige digitalisering af dokumentationen, hvis arkiveringsversioner skal oparbejdes til 2004-regler. Herudover kan det være ønskeligt med udarbejdelse af yderligere dokumentation. Som påpeget i forbindelse med den ideelle liste er det centralt for en fortolkning af data til information, at der findes en arkivrettet dokumentation af, hvordan arkiveringsversionens datafragmenter skal samles til dataelementer, svarende til det oprindelige systems skærbilleder og udskrifter. Der vil måske nok i de ældre afleveringer være en dokumentation af det oprindelige systems skærbilleder med videre, men nok ikke en dokumentation der knytter an til arkiveringsversionen. Sådan er det, og det er der næppe så meget at gøre ved. Hvis arkiveringsversionen stammer fra et nyere og kompliceret system, hvor det vurderes at være et problem fx i forhold til arkivets dokumentationsvirksomhed, kan det overvejes, om det kan lade sig gøre at skaffe de nødvendige oplysninger.

Det skal dog pointeres som ret væsentligt, at problemerne med manglende dokumentation af sammenkoblingen af datafragmenter til dataelementer næppe vil være så store som tidligere skitseret, når det gælder afleveringer fra IT-systemer, der er ældre end 1990 eller måske endda lidt senere. Disse systemer var af en anden type end dem, vi kender i dag, hvor fleksibiliteten i visningen af data er stor. I de ældre systemer er den struktur, som man fik data vist eller typisk udskrevet i, meget tættere på den struktur, som data fysisk er lagret i.

Som foreslået i den ideelle liste kunne det også være ønskeligt med en arkivskabt dokumentation, som kunne bruges i tilknytning til en arkivdatabase, og som kunne anvendes af den forsker, for hvem det ikke er nødvendigt eller muligt at bruge megen tid på at sætte sig ind i den oprindelige dokumentation. Der er tidligere argumenteret for, at arkivaren i forbindelse med bevarings- og kassationsbehandling og afleveringssag vil samle en viden sammen, som så blot skal nedskrives i et struktureret dokument. På samme måde vil arbejdet med oparbejdning til 2004-standard og konverteringer give de implicerede medarbejdere en viden, der vil kunne bevares for eftertiden ved hjælp af et arkivskabt dokument.

⁷³ Der vil som tidligere nævnt formodentlig være nogle afleveringer i perioden op til 1998, hvor forskellene til afleveringer efter 1998-regler vil være begrænsede. Forskellene bliver naturligvis større, jo længere bagud i tid man kommer.

7. AFRUNDING

Som grundlag for konkrete forslag til dokumentation af data i systemuafhængige arkiveringsversioner er der lavet en informationsvidenskabelig og arkivteoretisk undersøgelse af arkivalier. Den informationsvidenskabelige del af undersøgelsen bygger på Börje Langefors' infologi-teori, hvor det centrale er, at data er tegn, der repræsenterer information, men informationen opstår først, når data fortolkes af en person. I fortolkningsprocessen indgår som en forudsætning fortolkerens forudviden, hvoraf en del vil være fælles med andre i et kommunikationsfællesskab. Den arkivteoretiske del har fokus på begreberne arkivalie, proveniens og autenticitet, og den er hovedsagelig baseret på en række udenlandske artikler fra de sidste femten år.

Et arkivalie uanset medie opstår, når data overføres til arkiv fra det oprindelige kommunikationsfællesskab, det vil sige fra den organisation, hvor data er opstået og anvendt. Hermed løsrives data fra kommunikationsfællesskabet, hvis fælles forudviden har været en af forudsætningerne for at fortolke data. Det er derfor væsentligt at få dokumenteret oprindelsen eller proveniensen. Det gælder ikke mindst e-arkivalier, hvor der i stadig flere tilfælde er tale om multiproveniensen, hvor oprindelsen og brugen af data kan fortabe sig, når data løsrives fra IT-systemet og kommunikationsfællesskabet.

For arkivalier på papir gælder det, at struktur og indhold er udelelige også ved overførsel til arkiv, således at dataelementer afleveres samlet, hvor dataelementet er den mindste enhed, som er nødvendig for at kunne fortolke data til information. For e-arkivalier gælder det, at indhold og struktur adskilles ved overførsel til en systemuafhængig arkiveringsversion, hvor struktur skal forstås som IT-systemet. Databasestrukturen bevares, evt. i en tilpasset version, men den indeholder kun datafragmenter. Koblingen af datafragmenterne til dataelementer foregår på brugergrænsefladeniveau, som ikke bevares.

Spørgsmålet om autenticitet er rejst, men langt fra besvaret, for hvordan skal begrebet forstås generelt i arkivsammenhæng, og hvordan skal det forstås i forhold til e-arkivalier? Er det tilstrækkeligt at dokumentere autenticiteten af tegn og datafragmenter, men ikke dataelementerne, der er grundlaget for fortolkning til information?

På baggrund af de teoretiske undersøgelser og inspireret af ISAD(G) og dataarkiver er der opstillet en liste over den ideelle dokumentation, som omfatter proveniens, indhold, struktur samt arkivskabte ændringer. Dokumentationen kan for det første være oprindelig, det vil sige arkivalier skabt til brug i det oprindelige kommunikationsfællesskab. Dokumentationen kan for det andet være skabt med henblik på arkiv for at dokumentere autenticiteten af data i arkiveringsversionen og for at muliggøre en fortolkning af arkiveringsversionens data ved en sammenkobling af datafragmenter til dataelementer. For det tredje kan dokumentationen være arkivskabt med det formål at lette arkivets egen og forskernes identificering og brug af arkiveringsversioner.

Det er et skøn, hvor væsentlige de enkelte dokumentationselementer på den ideelle liste er, og dele af dokumentationen kan være bevaret andetsteds. Jo nærmere på det oprindelige IT-systems udformning og brug dokumentationen er, jo mere sandsynligt er det, at den ikke bevares andetsteds. Det grundlæggende er sammenkoblingen af datafragmenter til dataelementer, det vil sige den sammenkobling af databasens felter, der i kombinati-

on med funktionalitet, fx beregninger, er sket på skærmen og i udskrifter. Det er væsentligt at sikre dokumentation af dette i tilknytning til arkiveringsversionen, da det uden denne dokumentation i mange tilfælde vil være ressourcekrævende og behæftet med stor usikkerhed at skabe dataelementer, der kan betragtes som autentisk repræsentation af information.

Litteraturliste

Electronic Records: A Workbook for Archivists – Draft (ICA, 2004)

Erlandsson, Alf, Electronic Records Management: A Literature Review (ICA Studies 10, 1996)

Guide for Managing Electronic Records from an Archival Perspective, ICA, Committee on Electronic Records (ICA Studies 8, 1997)

Hansen, Per H. og Jeppe Nevers (red.), Historiefagets teoretiske udfordring (2004)

Information og dokumentation – Dokumentstyring, DS/ISO 15489-1

ISAD(G): General International Standard Archival Description, second edition (ICA, 1999)

ISAAR (CPF): International Standard Archival Authority Record for Corporate Bodies, Persons and Families, ICA (2004)

Jørgensen, Sebastian Olden, Til kilderne! (Kbh., 2001)

Kassation og bevaring af arkivalier. En intern vejledning (Rigsarkivet, 1979)

Kjeldgaard, Anne Sofie Fink, ”Bevaring af viden – Om oparbejdning af data i Dansk Data Arkiv”, Metode og Data, nr. 89, 2003

Langefors, Börje and Kjell Samuelson, Information and Data in Systems (Petrocelli, New York, 1976)

Langefors, Börje, Hermeneutics, Infology and Information Systems, Forskningsrapport nr. 77-10 (Informationsforskningsafdelingen, Handelshøjskolen i København)

Langefors, Börje, Theoretical Analysis of Information Systems (Lund, 1978)

Langefors, Börje, Kartläggning av informationsbehov i Malmberg, Anders, Decentraliseringen och den ekonomiska uppföljningen i foretaget (Stockholm, 1981)

Langefors, Börje, Essays on Infology (Studentlitteratur, 1995)

The Long-term Preservation of Authentic Electronic Records: Findings of the InterPA-RES Project (<http://www.inter pares.org>)

Nielsen, Anders Bo, ”Elektroniske arkivaliers autenticitet”, Tidsskriftet Arkiv, nr. 6, 2001, s. 26 - 32

Rasmussen, K. Boye, Datadokumentation (2000)

Reference Model for an Open Archival Information System (OAIS), Consultative Committee for Space Data Systems (2002)
(<http://ssdoo.gsfc.nasa.gov/nost/isoas/overview.html>)

Rothenberg, Jeff, Avoiding technological Quicksand (1999)

Schellenberg, T.R., The Appraisal of Modern Public Records (Bulletins of The National Archives, 1956)

Schellenberg, T.R., The Management of Archives (1965)

Schürer, K, Better access to electronic information to the citizen, INSAR, Supplement V (European Communities, 2001)

Shepherd, Elizabeth and Charlotte Smith, "The Application of ISAD(G) to the Description of Archival Datasets", Journal of the Society of Archivists, Vol. 21, No. 1, 2000

The Concept of Record, Second Stockholm Conference on Archival Science and the Concept of Record 30-31 May 1996 (1998)

The Principle of Provenance, First Stockholm Conference on Archival Theory and the Principle of Provenance 2-3 sept 1993 (1994)