



IT- og Telestyrelsen
Ministeriet for Videnskab
Teknologi og Udvikling

FESD Skanningsmodul Version 2.0

Høring

Dette udkast til forslag til FESD-standard er i offentlig høring i perioden fra 12. juli 2007 til 22 august 2007.

IT- og Telestyrelsen København den 12. juli 2007

FESD-standardisering
Skanningsmodul. Modul Version (2.0)

Kolofon:

FESD-standardisering. Skanningsmodul. Modul Version (2.0)

Denne standard kan frit anvendes af alle. Citeres der fra standarden i andre publikationer til offentligheden, skal der angives korrekt kildehenvisning.

Forslag til FESD standarder udarbejdes af IT- og Telestyrelsen, IT-Arkitektur kontoret, FESD standardiseringsgruppen i samarbejde med de tre FESD leverandører Software Innovation A/S, Accenture I/S og CSC Danmark A/S.

Kontaktperson i FESD-standardisering:

Projektleder Palle Aagaard, Mail-adresse paa@itst.dk

Telefon 33 37 92 95 (direkte)

Accenture I/S

Lautrupsgade 7

2100 København Ø

Telefon: 72 28 80 00

Web-adresse: <http://www.accenture-fesd.dk/>

CSC Danmark A/S

Retortvej 8

1780 København V

Telefon: 36 14 40 00

Web-adresse: <http://www.fesd-alliancen.dk/>

Software Innovation A/S

Nærum Hovedgade 10

DK-2850 Nærum

Telefon: 45 58 88 88

Web-adresse: <http://www.softwareinnovation.dk/>

Ministeriet for Videnskab, Teknologi og Udvikling

IT- og Telestyrelsen

IT-Arkitektur kontoret

National IT and Telecom Agency

Ministry of Science, Technology and Innovation

Holsteinsgade 63

DK-2100 København Ø

Telf. +45 35 45 00 00

Fax. +45 35 45 00 10

<http://www.itst.dk>

itst@itst.dk

Indholdsfortegnelse

| | | |
|--|---|-----------|
| 1 | FORORD | 5 |
| 2 | DEL A – SKANNINGSMODUL | 7 |
| 2.1 | Indledning | 7 |
| 2.2 | Afgrænsning..... | 8 |
| 2.3 | Skanningsmodulet i en større sammenhæng | 8 |
| 2.4 | Beskrivelse af skanningsmodulet | 9 |
| 3 | DEL B – FUNKTIONER I SKANNINGSMODUL | 11 |
| 4 | DEL C – INTERFACES | 13 |
| 4.1 | FESDSkanningsinterface..... | 13 |
| 4.2 | Myndigheds- eller sektorspecifikke udvidelser af skemaet..... | 14 |
| 4.3 | Afledte krav til kernesystemet | 24 |
| 5 | BILAG A | 25 |
| Anvendte typer i FESD-modellerne..... | | 25 |
| | integer..... | 25 |
| | boolean | 25 |
| | Identifikationstyper | 26 |
| | char | 26 |
| | date, dateTime og time. | 26 |
| | string..... | 27 |
| | float..... | 27 |

1 Forord

Den offentlige sektors IT-systemer på statsligt, kommunalt og regionalt niveau skal kunne spille sikkert og effektivt sammen. Derfor arbejdes der målrettet på at få gennemført fælles standarder for elektronisk sags- og dokumenthåndtering - den såkaldte FESD-standard. Målet med standardiseringsarbejdet er at fremme digital forvaltning i den offentlige sektor, og midlet er at sikre, at de forskellige elektroniske sags- og dokumenthåndteringssystemer (ESDH) får en fælles kernefunktionalitet, og at det samtidig sikres, at denne kerne videreudvikles ensartet. En fælles kernefunktionalitet skal sikre:

- at der kan foretages sagsbehandling på tværs af flere organisationer
- at myndigheder, der arbejder med åbne sager, kan lægges sammen
- at der kan flyttes opgaver mellem forskellige myndigheder

I forlængelse af FESD-projektkonkurrencen, som havde sin afslutning primo 2004, og hvor der blev fundet tre FESD-leverandører, blev det i forbindelse med kontraktforhandlingerne besluttet at starte en standardiseringsproces – den såkaldte FESD-standardisering.

For at sikre interoperabiliteten, både til andre systemer, men også så tredjepart kan udvikle moduler til systemet, blev det anset for afgørende, at der udvikles en fælles offentlig datamodel samt andre standarder på ESDH-området.

Koordinering af FESD-standardiseringen er efterfølgende lagt i IT- og Telestyrelsen (ITST). Den konkrete udarbejdelse af forslag/udkast til standarder foregår i et samarbejde mellem de tre FESD-leverandører og en FESD-standardiseringsgruppe i ITST.

Arbejdet med forslag/udkast til standarder tager udgangspunkt i Noark 4's datamodel og databeskrivelser samt leverandørernes løsninger. Standarderne kan afvige fra Noark 4 på de områder, hvor det er nødvendigt for at understøtte dansk forvaltningspraksis, eller hvor parterne i FESD-standardiseringen kan opnå enighed om en afvigelse.

Udkast/forslag sendes herefter i offentlig høring i ca. 1 måned. FESD-standardiseringsgruppen tilretter og færdiggør på baggrund af høringen de endelige "Forslag til standarder".

Standardforslagene forelægges herefter OIO-Datastandardiseringskomiteen til godkendelse.

Efter den samlede godkendelse bliver standarderne således offentliggjort og indgår i IT- og Telestyrelsens "OIO-Katalog", som indeholder en oversigt over godkendte og anbefalede standarder til digital forvaltning i det offentlige.

I standarden kan forekomme brug af særligt ordvalg. Følgende termer anvendes konsekvent i den følgende betydning:

- **"skal"/"obligatorisk"**: betyder, at den nævnte metode/element/mulighed/etc. skal benyttes eller skal forefindes – dvs. må ikke udelades.
- **"må ikke"**: betyder, at den nævnte metode/element/mulighed/etc. ikke må forefindes eller må ikke benyttes.
- **"bør"/"anbefalet"**: betyder, at det i høj grad anbefales, at den nævnte metode/element/mulighed/etc. benyttes eller forefindes. Der skal være tungtvejende grunde til at udelade.
- **"kan"/"optional"**: betyder, at den nævnte metode/element/mulighed/etc. er en valgmulighed og derfor valgfri at medtage.

Grundlaget for FESD-datamodellen er blevet udarbejdet på en periode på mere end 2 år, hvor der er udarbejdet standarder for de forskellige delområder. Arbejdsmetoder, terminologi og anvendelse af datatyper har ændret sig i denne periode – FESD-standardiseringsgruppen har f.eks. indført en konsekvent brug af

UML-notation i de senere standarder. Konsolideringen af datamodellen har derfor forudsat, at der blev defineret en fælles modelleringsmetode og et sæt af primitive datatyper, der var kompatibelt med alle del-datamodellerne. De primitive datatyper, som modellen er opbygget af, fremgår af bilag A.

2 Del A – Skanningsmodul

Dette er version 2.0 af standarden ”FESD skanningsmodul”. Første udgave af denne standard var version 1.0 af 8 september 2005.

Ændringer fra version 1.0 til version 2.0 er som følger:

Ændringerne til version 2.0, er alene en udvidelse af det skema som benyttes til udveksling af data mellem skanningsmodul og FESD ESDH systemet – med andre ord, udvidelse af de muligheder, skanningsmodul har for at sende metadata om de indskannede dokumenter til FESD ESDH systemet. Ændringerne afspejler de erfaringer der er opsamlet fra de implementeringer der er foretaget af standarden. Der er i forbindelse med revisionen, gennemført høringsmøde med en række leverandører der benytter standarden i forbindelse med deres produkter.

Udover en række specifikke metadata elementer, er standarden udvidet med en mulighed for at beskrive myndigheds- og/eller sektorspecifikke metadata. Grænsefladen i standarden er således blevet betydelig mere fleksibel i forhold til evt. udvidelser af skanningsmodulets tolkningsmuligheder i de dokumenter der indskannes.

Konkret er der tilføjet følgende elementer til skemaet: Emnereference, reference til arkiv, afsenderoplysninger (adresse), kundespecifik udvidelse i forhold til metadata om dokumentet, Kundenspecifik udvidelse i forhold til dokumentindhold (blanket). Der er endvidere medtaget mulighed for at varianter af dokumenter, i form af OCR behandlede udgaver kan medtages i udvekslingen.

Andre ændringer er først og fremmest sket foranlediget af det konsolideringsarbejde, som er udført af FESD standardiseringen. Dette konsolideringsarbejde har været nødvendiggjort af, at FESD standarder udformes projekt-orienteret, dvs. de enkelte dele af den samlede FESD standard udformes i ”pakker” over en årrække. Dette betyder, at der må foregå en iterativ proces - kaldet konsolidering – for at få samjusteret og evt. konsekvensrettet. Ved den nævnte projektmæssige arbejdsform kan der nemlig især opstå den vanskelighed, at oprindeligt standardiserede ”pakker” har mangler i forhold til behov der opdages ved senere standarder. Netop sådanne justeringer i de enkelte standarder er derfor primært af redaktionel karakter, idet de bedst kan karakteriseres som konsekvensrettelser foranlediget af senere godkendte standard-pakker.

Alle attributter, der er unikke ID’er, foreskrives at skulle udfyldes med UUID’er jf. de seneste FESD standarder. UUID’er bliver således fælles for alle FESD standarder.

I datamodellen forekommer attributter nu både på dansk og engelsk. Ligeledes er der tilføjet et – for alle FESD standarder – nyt fælles forord samt anvendelse fornyede designretningslinier for FESD standarderne.

2.1 Indledning

Modul i FESD standardisering betragtes overordnet som ”en veldefineret opgave med et fast defineret sæt af funktioner og grænseflader”, der enten er ”programmeret som en selvstændig del af løsningen, der kan fungerer selvstændigt og/eller sammen med løsningens øvrige dele” eller ”er et veldefineret sæt af funktioner, der kan frakobles, og hvis funktioner kan overtages af eksterne systemer ved brug af veldefinerede grænseflader”.

Udarbejdelsen af et selvstændigt skanningsmodul skal, udover den overordnede vision om effektivisering og FESD-behovsopgørelsens arkitekturprincip om modularisering, sikre opfyldelsen af visionen om, at: ”Understøtte samarbejde og omstrukturering mellem myndigheder”. Der er i den offentlige forvaltning udbredt behov for i praksis og effektivt at kunne gennemføre omlægning af opgaver mellem myndigheder, og dette behov er forstærket af brede strukturomlægninger i de kommende år. Det indebærer, at et skanningsmodul skal kunne håndtere ændrede krav til organisationen fx i forbindelse med:

Væsentligt øgede eller reducerede mængder af materiale der skal skannes:

Behov for større eller mindre skannere

Behov for at ændre i arbejdsgange omkring skanning

Behov for at ændre antallet af fysiske lokaliteter

Ændret opgavefordeling:

Behov for at skanne andre formater

Behov for at skanne i farve

Behov for skanning og behandling af skemaer

Ændrede behov for kvalitetssikring

Ændrede behov for metadata

Det er også en væsentlig forudsætning for standardiseringen af skanningsmodulet, at den beskrevne grænseflade og funktionalitet åbner for, at 3. parts leverandører leverer skanningsmoduler til FESD-kompatible ESDH-løsninger. Standarden sikrer, at skanning til FESD-kompatible ESDH-systemer kan tilbydes som en selvstændig serviceydelse (outsourcing).

2.2 Afgrænsning

Standarden for skanningsmodul har sit udgangspunkt i FESD-rammekontrakten bilag 10 Tillægsydelser afsnit 10.4 Scanning og OCR-facilitet. Standarden baserer sig ikke på eller har sit udgangspunkt i den norske NOARK-4 standard som det er gældende for andre FESD standardiserings arbejde, da NOARK ikke opererer med et skanningsmodul.

2.3 Skanningsmodulet i en større sammenhæng

For bedre at kunne forstå og vurdere de efterfølgende krav til skanningsmodulet, beskrives et eksempel på en indskanningsproces med kommentarer. Indskanningsprocessen kan meget overordnet beskrives således:

1. Det/de ønskede dokument(er) gøres klar til indskanning (det omfatter fx, at dokumenterne tages ud af konvolutter, foldes ud, clips og hæfteklammer fjernes, og dokumenterne lægges i arkføderen på skanneren. Dette er en ren manuel proces).
2. Eventuelt kan de enkelte dokumenter adskilles ved hjælp af skilleblade, eventuelt påført en række metadata fx i form af strekkoder, ligesom hoveddokumentet eventuelt adskilles fra eventuelle bilag.
3. Det/de ønskede dokument(er) indskannes.
4. Der kan eventuelt foretages en OCR-behandling af det indskannede, eller denne behandling kan foretages på et senere tidspunkt i processen. Da OCR-behandling typisk er en meget tidskrævende proces, vil den typisk ske senere for ikke at forsinke den videre fordeling af de indskannede dokumenter.
5. Der kan foretages en optisk kvalitetskontrol af det indskannede: Er der det rigtige antal sider, og er alle siderne læselige.
6. Eventuelt genindskannes dokumenter, hvor kvaliteten ikke er tilfredsstillende.
7. Der kan ske en 'forjournalisering', hvor en række metadata om dokumentet påføres. Visse metadata påføres automatisk og må ikke kunne ændres senere (fx skanningsdato, skanningstidspunkt, skanningsstation og initialer for vedkommende, som har foretaget indskanningen), og en række andre metadata kan være blevet udfyldt via strekkoder på skilleark.
8. Dokumentet fordeles.

Standardiseringen må ikke lægge bånd på organiseringen af arbejdet omkring indskanning. Der er en række funktionaliteter, som skal findes både i skanningsmodulet og i kernesystemet. Hermed sikres det, at den enkelte organisation frit kan vælge indskanningsprocedure. Som et eksempel skal den enkelte organisation frit kunne vælge,

- om kvalitetssikringen skal ske i skanningsmodulet eller i kernesystemet

- hvor meget af (for)journaliseringen der skal foretages i skanningsmodulet

Skanningsmodulet skal – udover at kunne håndtere indskanningen af de (papir)dokumenter som ønskes gjort tilgængelige i elektronisk form – også kunne understøtte kvalitetskontrollen og den efterfølgende genindskanning samt udfyldelsen af en række metadata om dokumentet, herunder eventuelt tilknytning til sag. Det betyder, at inden skanningsmodulet 'sender' det indskannede dokument til ESDH-systemet, skal skanningsmodulet stille funktionalitet til rådighed til både at udfylde en række metadata via stregkoder på skilleblade henholdsvis via adgang til manuel påføring af metadata.

Indskanningsproceduren skal kunne tilrettelægges således, at den er effektiv og kan understøtte de særlige behov, de enkelte organisationer har, og så skal den sikre den størst mulige retslige bevisværdi af de digitale dokumenter. Her henvises der til IT-Sikkerhedsrådets publikation ”Digitale dokumenters bevisværdi - Introduktion og vejledning” fra 27. januar 1999, som også er en del af FESD krav 2.4.1.

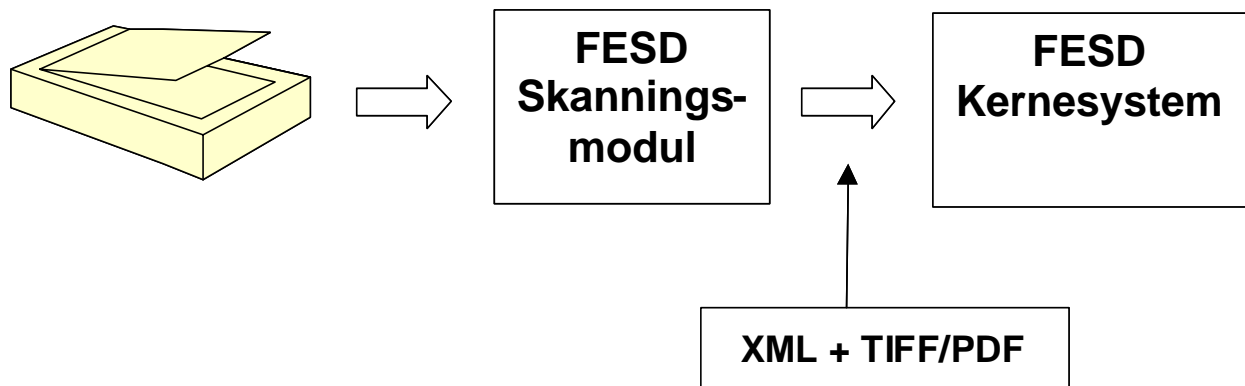
2.4 Beskrivelse af skanningsmodulet

Skanningsmodulet består af en eller flere softwarekomponenter, der tilsammen implementerer skanningsfunktionalitet til brug for ESDH. Skanningsmodulet har den fysiske skanner som input, typisk i form af en eller flere TIFF-filer, og som output afleveres et indskannet dokument i enten TIFF eller PDF samt en XML-fil indeholdende metadata (Se detaljeret beskrivelse del C). Kravet til ESDH-kernesystemet vil derfor være, at ESDH-systemet skal kunne modtage XML-opmærkede dokumenter, som beskrevet efterfølgende i dette dokument.

Der stilles ingen eksplicitte krav til TIFF- eller PDF-formater ved indskanning.

Det kan være en fordel at benytte samme format, som kræves af Statens Arkiver ved aflevering, men det er ingen betingelse.

Nedenstående tegning tjener udelukkende til konceptuel forståelse af kravene. Der kan derfor ikke udledes obligatoriske krav af tegningen.



Figur 1: Beskrivelse af skanningsproces

Standardiseringen af skanningsmodulet tager udgangspunkt i, at der eksisterer en række metadata om dokumentet, som man ønsker at registrere i sammenhæng med skanningsprocessen. Eksempler på berigende metadata kan være Dokumenttitel, Sagsnummer, Dokumenttype etc., mens fysiske metadata kan være Skanningsoperatør, Skanningsstation, Fordelingsinformation etc.

Standardisering af skanningsmodulet har som formål at sikre størst mulig fleksibilitet for den praktiske tilrettelæggelse af arbejdsgange i de enkelte myndigheder.

For skanningsmodulet betyder det konkret, at det er fleksibelt med hensyn til:

- Hvilke metadata der kan påføres dokumentet i forbindelse med skanningen – visse organisationer vil eventuelt påføre mange metadata, mens andre vil påføre meget få metadata. Dette har betydning for organisering af arbejdet med hensyn til hvilke opgaver, der skal varetages af henholdsvis støttefunktioner og sagsbehandlere.
- Hvornår metadata påføres skannede dokumenter – i forbindelse med skanning eller i afdelingen hos den enkelte medarbejder.
- Hvornår og hvordan de skannede dokumenter kvalitetssikres i arbejdsprocessen.

Samlet set bør standarden ikke sætte begrænsninger for, hvordan arbejdet med skanning, journalisering og distribution af dokumenter organiseres.

Skanningsmodulet skal opfylde en række funktionelle krav, som er beskrevet i det efterfølgende.

3 Del B – Funktioner i skanningsmodul

Herunder er beskrevet de funktionelle krav, som skanningsmodulet skal kunne understøtte. Kravene er nummeret S1 til S18. Under Obligatorisk er angivet O, hvis funktionen er obligatorisk, og A hvis funktionen er anbefalet.

| Funktion | Obligatorisk | Beskrivelse |
|----------|--------------|---|
| S1 | O | Modulet skal kunne aflevere dokumenter til ESDH-systemet i formatet TIFF. Som minimum skal modulet kunne aflevere TIFF i en version som beskrevet i Statens Arkivers afleveringskrav. |
| S2 | A | Modulet skal kunne aflevere dokumenter til ESDH-systemet i formatet PDF. |
| S3 | O | Modulet skal kunne aflevere dokumenter til ESDH-systemet ved hjælp af input interfacet beskrevet under afsnittet "Interfaces", der er beskrevet i afsnittet om interfaces. |
| S4 | O | Modulet skal både kunne håndtere indskanning af enkelt dokumenter og håndtere indskanning af flere samtidige dokumenter og opdele disse i enkelt dokumenter. Hvert dokument kan indeholde en eller flere sider. |
| S5 | O | Modulet skal kunne håndtere et indskannet dokument, som er opdelt i hoveddokument og et eller flere underdokumenter. Underdokumenterne (bilagene) skal opfattes som selvstændige dokumenter (egen dokumentidentifikation). |
| S6 | A | Modulet skal kunne understøtte skannere via interfacene ISIS og TWAIN. |
| S7 | A | Modulet skal kunne understøtte netværksskannere, som afleverer de indskannede dokumenter i filsystem. |
| S8 | O | Modulet skal give mulighed for, at der for hvert enkelt dokument kan påføres tilhørende metadata, svarende til de muligheder, som input interfacet beskrevet under afsnittet "Interfaces" tilbyder. |
| S9 | O | Modulet skal automatisk påføre skanningsoperatør, skanningsstation og tidsstempel som metadata i XML-opmærkningsfilen. |
| S10 | O | Modulet skal kunne påføre metadata enten via strekkodeskilleark eller via OCR genkendelse af dokumenterne. Der er ingen eksplicitte krav til typen af strekkoder, som modulet kan genkende, ligesom der ikke er krav til format og opbygning af strekkoden. |
| S11 | O | Modulet skal kunne påføre metadata til de indskannede dokumenter ved direkte indtastning af metadata. |
| S12 | O | Modulet skal kunne håndtere indskanning af dokumenter i sort-hvid, gråtoner og farve. |
| S13 | O | Modulet skal kunne håndtere indskanning af dokumenter i A5 til A0 format. |
| S14 | O | Modulet skal give mulighed for, at der foretages en kvalitetskontrol af det indskannede, før det afleveres til ESDH-systemet. Kvalitetskontrol skal forstås som en optisk inspektion af det indskannede. |
| S15 | O | Modulet skal kunne zoome ind og ud på det indskannede dokument, som led i kvalitetskontrollen. |

| | | |
|-----|---|---|
| S16 | O | Modulet skal give mulighed for, at der foretages opretning af skæve sider, vending af sider, kontrastforbedring samt frasortering af tomme sider i forbindelse med kvalitetskontrollen. |
| S17 | O | Modulet skal give mulighed for i forbindelse med kvalitetskontrol at afvise et indskannet dokument, før det afleveres til ESDH-systemet. |
| S18 | A | Modulet skal kunne påføre de indskannede dokumenter et sæt default metadata, hvis ikke der kan påføres dokumentspecifikke metadata via strekkoder, OCR eller direkte indtastning. |
| S19 | O | Modulet må ikke påføre nogen begrænsning i antallet af skannere som kan tilsluttes ESDH-systemet eller i antallet af dokumenter som kan indskannes og lagres i ESDH-systemet, andet end de begrænsninger som stammer fra hardware eller systemsoftware. |
| S20 | O | Modulet skal kunne indskanne dokumenter med tekst på både for- og bagside. |
| S21 | O | Modulet skal kunne understøtte genindskanning af dokumenter som er blevet afvist i kvalitetskontrollen. |

4 Del C – Interfaces

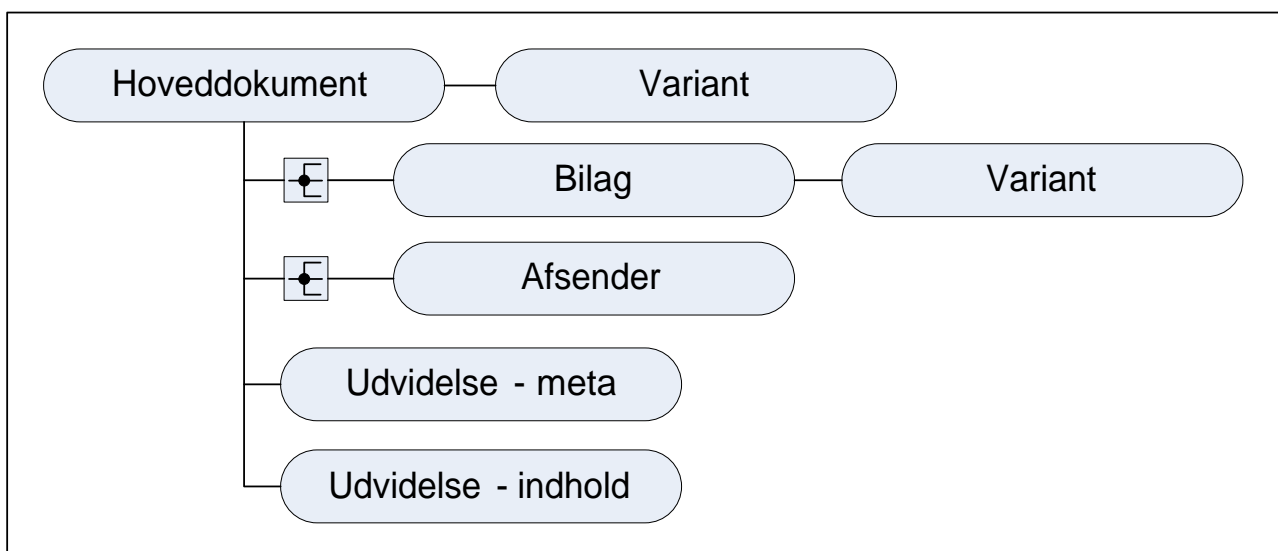
Der findes to interfaces: Et input interface til skanningsmodulet og et output interface fra skanningsmodulet, svarende til et input interface til ESDH systemet. Input interfacet til skanningsmodulet er det interface, hvor skanneren afleverer det indskannede dokument til skanningsmodulet. Denne standardiseringsbeskrivelse beskæftiger sig ikke med dette interface, men udelukkende med det andet interface mod ESDH-systemet, som herefter betegnes FESDSkanningsinterfacet.

4.1 FESDSkanningsinterface

FESDSkanningsinterfacet behandler én dokumentpakke af gangen. En dokumentpakke består af:

- Et hoveddokument (en TIFF- eller PDF-fil)
- 0, 1 eller flere bilag (en TIFF- eller PDF-fil pr. bilag)
- 0 eller 1 variant pr. hoveddokument/bilag
- Én XML-fil

Strukturen i XML-dokumentet, for et enkelt indskannet dokument kan illustreres således:



Dokumentvarianter

I forbindelse med en evt. OCR tolkning af de indskannede dokumenter, vil man typisk have brug for at kunne sende det indskannede dokument over til ESDH-systemet, både i et visningsformat og i et OCR behandlet format.

Dette realiseres ved anvendelse af variantbegrebet, hvor såvel hoveddokument, som bilag kan medsendes i to udgaver. Varianten skal opfattes som en tolket udgave af enten hoveddokument og bilag, og kan dermed i det modtagende ESDH-system, indekseres og anvendes i forbindelse med fritekstsøgning.

Hver dokumentpakke afleveres fra skanningsmodulet til ESDH-systemet via filsystemet, således at der oprettes et filkatalog for hver dokumentpakke, hvori alle dokumentpakkens filer placeres. Filkataloget skal navngives efter følgende syntax:

"FESDScan"<Skanningsstation>"-"<Skandato><Skantid>

Hvor

- <Skanningsstation> er en fri alfanumerisk identifikation af skanningsstationen på mellem 1 og 50 tegn.
- <Skandato> er datoen for indskanningen på formen CCYYDDMM.
- <Skantid> er tidspunktet for indskanningen på formen HHMMSS.

Det er ESDH-systemet ansvar at slette katalogerne inklusiv filerne, efter dokumenterne er registeret i ESDH-systemet.

Hoveddokumenter og bilag kan frit navngives i filsystemet af skanningsmodulet. Identifikationen af hoveddokument og bilag sker gennem XML-elementerne <FileName> og <ScannedDocument> i den tilhørende XML-fil. XML-filen skal altid have navnet "FESDScan.xml". XML-filen skal overholde skemaet:

<http://rep.oio.dk/fesd.dk/xml/schemas/2005/02/01/FESDScan.xsd>.

Dette skema kan ses efter følgende oversigt over særlige skanningsmetadata herunder:

4.2 Myndigheds- eller sektorspecifikke udvidelser af skemaet

Datafølgelsesdelen indeholder generelle metadata for sager, dokumenter og parter. Normalt vil disse være dækkende i forhold til at beskrive de sager og dokumenter der oversendes, men der kan være behov for sektorspecifikke eller myndighedsspecifikke metadata, som vil kunne berige de oversendte sager/dokumenter og dermed potentielt berige en given sagsproces.

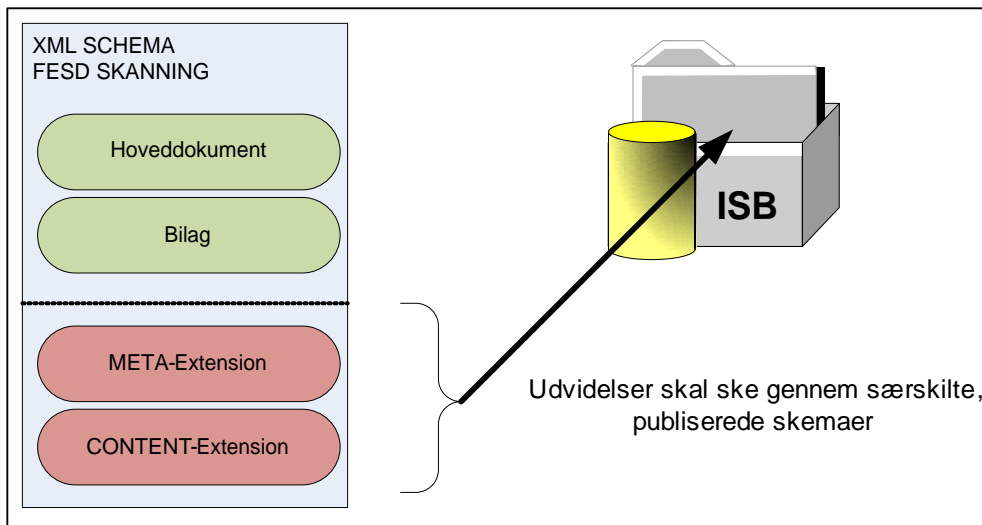
Generelt har man i standardiseringsarbejdet været tilbageholdende med at tillade denne type udvidelser, ud fra den generelle opfattelse, at individualiseringen vil svække skemaerne og dermed standarden. NDR Reglerne er således også særdeles restriktiv på dette område. På den anden side, må det konstateres at der er adækvate behov for udvidelser, af en karakter som ikke har bred interesse og som dermed kan indeholdes i standarden.

Datafølgelsesdelen indeholder derfor en mulighed for udvidelse, som kan benyttes. Udvidelsesmuligheden kan benyttes indenfor de nedenfor beskrevne restriktioner.

Anvendelse af myndigheds- eller sektorspecifikke udvidelser af skemaet

De forskellige typer af dokumenter hhv. typer af myndigheder giver forskellige muligheder for at berige dokumenterne i forbindelse med indskanningen. Eksempelvis vil man på det kommunale område kunne aflæse en stregkode på de standardiserede blanketter og dermed tilvejebringe information om emne, samt have mulighed for intelligent aflæsning af informationer i blanketten.

Overordnet kan den påtænkte udvidelsesmulighed illustreres på følgende måde:



De to udvidelsesformer dækker:

- 1) Beskrivende metadata om dokumentet
Yderligere, beskrivende attributter om dokumentet
- 2) Aflæst indhold fra dokumentet.
Dokumentindhold – struktureret – tolket fra f.eks. en blanket

Tilføjelse af de to typer af udvidelser, realiseres gennem et såkaldt ANY element, der som bekendt kan benyttes uden restriktioner i XML skemaet. En fri modellering af ANY elementet, vil svække standarden, og elementet vil således kun kunne benyttes indenfor standarden, såfremt en række betingelser er overholdt:

- 1) Udvidelser skal realiseres gennem særskilte skemaer, som er publiceret på ISB'en
- 2) Skemaerne skal følge NDR reglerne i forhold til udformning, mv.
- 3) Skemaerne skal formålsbeskrives, og må ikke dække generelle forhold som rettelig skal adresseres i standarden og ikke som en udvidelse.
- 4) Der skal være reference til ansvarshavende for udvidelsen

Kravet om publicering af udvidelserne betyder at modtageren vil være i stand til at validere de udvidelser der er foretaget. Det vil også betyde, at standardiseringsenheden vil kunne opsamle og vurdere de forskellige udvidelser der foretages, med henblik på at revidere standarden, og evt. tilføje yderligere, generelle metadata til standarden.

Beskrivelse af metadata for elementet ScannedDocument

| Attribut Dan | Attribut Eng | Type | Kardinalitet | Beskrivelse | Mapning til NOARK4 |
|-----------------------------|------------------------|----------|--------------|--|-----------------------|
| emneIdentifikator | termIdentifier | char(34) | [0..1] | <p>EmneID kan benyttes til at angive den emnemæssige indplacering af dokumentet. EmneID er ikke den tekniske reference emnet i en given emnesystematik, jf. FESD datamodel vedrørende emnesystematik.</p> <p>Eksempelvis vil emneID, i forbindelse med anvendelse af KL's emnesystematik KLE, kunne have værdien "12.01.02", som angiver emnet til at være "</p> <p>Det er ikke en forudsætning at skanningsmodulet kender den pågældende emnesystematik, men skal alene formidle tilknytningen eller indtastet emne.</p> | |
| arkivIdentifikator | archiveIdentifier | char(34) | | Dette felt benyttes til at angive en reference til det arkiv, det indskannede dokument skal lagres i. For uddybning af begrebet arkiver og dets anvendelse, henvises til FESD Datamodel vedrørende Arkivstruktur. | |
| dokumentTypeForkortelse | documentTypeCode | Code | [1] | Dette felt kan lagres direkte i JournalPost feltet JournalPostDokumentTypeForkortelse. | |
| modtagerBrugerIdentifikator | recipientUserReference | Char(50) | [0..1] | Dette felt benyttes som forslag til at registrere feltet JournalPostAnsvarligSagsbehandler, når JournalPosten oprettes. ModtagerBrugerIdentifikator henviser til feltet BrugerNavn fra BrugerAdministration, og det er ESDH-systemets ansvar at udlede værdien af JournalPostAnsvarligSagsbehandler feltet ved en JournalPost oprettelse. Indtil oprettelsen af JournalPosten kan værdien bruges som fordelingsinformation. ESDH-systemet kan også benytte dette felt som forslag til SagAnsvarligSagsbehandler, hvis der oprettes en sag på baggrund af det indskannede dokument. | |
| modtagerEnhedIdentifikator | recipientUnitReference | Char(50) | [0..1] | Dette felt benyttes som forslag til at registrere feltet JournalPostAnsvarligEnhed, når JournalPosten oprettes. ModtagerEnhedIdentifikator henviser til feltet EnhedsNavn fra BrugerAdmi- | |

| Attribut Dan | Attribut Eng | Type | Kardinalitet | Beskrivelse | Mapning til NOARK4 |
|---------------------------------------|-------------------------|----------|--------------|---|-----------------------|
| | | | | nistration, og det er ESDH-systemets ansvar at udlede værdien af JournalPostAnsvarligEnhed feltet ved en JournalPost oprettelse. Indtil oprettelsen af JournalPosten kan værdien bruges som fordelingsinformation. ESDH-systemet kan også benytte dette felt som forslag til SagAnsvarligEnhed, hvis der oprettes en sag på baggrund af det indskannede dokument. | |
| skannerOperatør | scannerOperatorName | Char(50) | [1] | Skanneroperatør navn. Dette felt lagres direkte i JournalPost feltet JournalPostSkanOperatør. | |
| skannerStation | scannerStationName | Char(50) | [1] | Skanner navn. Dette felt lagres direkte i JournalPost feltet JournalPostSkanStation. | |
| skannerDatoTid | scanDateTime | DateTime | [1] | Dato og tidspunkt for skanning. Dette felt lagres opdelt i JournalPost felterne JournalPostSkanDato og JournalPostSkanTid. Formatet er CCYYDDMMThh:mm:ss, Jf. ISO 8601. | |
| foregaaendeJournal- PostHenvisning | previousRecordReference | LongCode | [0..1] | Dette felt udfyldes, hvis et dokument, som allerede er blevet journaliseret, skal indskannes igen (fx fordi den optiske kvalitetskontrol viser, at dokumentet er ulæseligt). Feltet udfyldes med journalpostnummeret på den journalpost, som skal erstattes af journaliseringen af det netop indskannede dokument. ESDH-systemet skal understøtte oprettelsen af en Kommentar på den gamle journalpost, samt skabe en relation mellem den gamle journalpost og den nye journalpost med en SagsReference relation, således at det tydeligt fremgår af registreringerne i ESDH-systemet at det nye dokument er en genindskanning af det gamle dokument. Dette felt skal altid være udfyldt ved genindskanning af et journaliseret dokument. | |

| Attribut Dan | Attribut Eng | Type | Kardinalitet | Beskrivelse | Mapning til NOARK4 |
|-------------------------------|----------------------------|---------------|--------------|---|-----------------------|
| dokumentFysiskDokumentNr | physicalDocumentIdentifier | ReferenceText | [0..1] | <p>Dette kan udfyldes med et fysisk dokument-nummer, fx et endorser-nummer fra skanneren. Benyttes til at identificere papirdokumentet udfra det indskannede elektroniske dokument i ESDH-systemet. Lagres i feltet DokumentFysiskDokumentNr.</p> <p>Dette felt i ESDH-systemet kan også benyttes ved anmodning om genindskanning af et allerede journaliseret dokument, da det entydigt identificerer papirdokumentet.</p> | |
| sagNrIdentifikator | caseFileNumberReference | LongCode | [0..1] | Dette felt benyttes som forslag til at registre feltet JournalPostSagID, når JournalPosten skal oprettes. SagNrIdentifikator henviser til feltet SagNr i FESD-datamodellen, og det er ESDH-systemets ansvar at udlede værdien af JournalPostSagID feltet ved en JournalPost oprettelse. | |
| dokumentTitelTekst | documentTitleText | Text | [0..1] | <p>Dokumenttitel.</p> <p>Dette felt kan lagres direkte i JournalPost feltet JournalPostIndholdsBeskrivelse.</p> | |
| dokumentDato | documentDate | Date | [0..1] | <p>Dokumentdato.</p> <p>Dette felt kan lagres direkte i JournalPost feltet JournalPostDokumentDato.</p> | |
| sikkerhedsKlassifikationsKode | securityClassificationCode | Code | [0..1] | Dette felt kan lagres direkte i JournalPost feltet JournalPostSikkerhedsKlassifikation. Sikkerhedsklassifikation er beskrevet i FESDstandarden for BrugerAdministration. | |
| adgangsGruppeHenvisning | accessGroupReference | Char(50) | [0..n] | Dette felt kan lagres direkte i JournalPost feltet JournalPostAdgangsGruppe. Der kan være flere adgangsgrupper knyttet til en JournalPost. Adgangsgrupper er beskrevet i FESDstandarden for BrugerAdministration. | |
| adgangsEnhed | accessUnitReference | Char(50) | [0..n] | Dette felt kan lagres direkte i JournalPost feltet JournalPostAdgangsEnhed. Der kan være flere adgangsenheder knyttet til en JournalPost. Adgangsenheder er beskrevet i FESDstandarden for BrugerAdministration. | |

| Attribut Dan | Attribut Eng | Type | Kardinalitet | Beskrivelse | Mapning til NOARK4 |
|-----------------|--------------------------|----------|--------------|--|-----------------------|
| filNavn | fileName | Char(50) | [1] | Dette element indeholder filnavnet for det tilhørende indskannede dokument. Dokumentet skal ligge i samme filkatalog som FESDSkan.xml | |
| skannetAppendix | scannedAppendixReference | Char(50) | [0..n] | Hvert bilag lagres som selvstændigt Dokument under JournalPosten. Se detaljeret beskrivelse nedenfor. Sammenknytningen af dokumenterne sker ved, at XML-filen udpeger både hoveddokumentet og bilagene ved deres filsystemnavne. | |

Beskrivelse af metadata for elementet ScannedDocumentSender

| Attribut Dan | Attribut Eng | Type | Kardinalitet | Beskrivelse | Mapning til NOARK4 |
|--------------------------------|-------------------------------|-------------|--------------|--|-----------------------|
| personCprNr | personCPRNumber | integer(10) | 0 -1 | Entydig identifikation på en person. Noter at cprNr ikke fungerer som entydig identifikator for klassen, da cprNr ikke kan forventes kendt for alle instanser. Udfyldes såfremt personOrganisationIndikator angiver at der er tale om en person. | |
| personAdresseringsNavn | personAddressName | Name | 1 | Fornavn, mellemnavn og efternavn, sammenstillet. Elementet er ikke nødvendigvis konstrueret af de selvstændige elementer fornavn, mellemnavn og efternavn, men kan i sin helhed være tolket eller indtastet i forbindelse med indskanningen. Udfyldes såfremt personOrganisationIndikator angiver at der er tale om en person. | |
| organisationCvrNr | organizationCVRNumber | integer(8) | 0-1 | Afsenderorganisationens CVR nummer. Udfyldes såfremt personOrganisationIndikator angiver at der er tale om en organisation. | |
| organisationOrganisations-Navn | organizationOrganization-Name | char(50) | 1 | Navnet på organisationen eller virksomheden. Udfyldes såfremt personOrganisationIndikator angiver at der er tale om en organisation. | |
| personOrganisationIndikator | personOrganizationIndicator | indicator | 1 | Angivelse af om afsender elementet er en person eller en organisation, og således hvilke oplysninger der skal være tilstede. | |

| Attribut Dan | Attribut Eng | Type | Kardinalitet | Beskrivelse | Mapning til NOARK4 |
|----------------------|--------------------------|----------|--------------|---|--------------------|
| adresseKommuneKode | addressMunicipalityCode | char(4) | 1 | Det af Indenrigsministeriet fastsatte nummer for kommunen. Refererer til klassen 'Kommune'. Herfra kan 'kommuneNavn' hentes. | |
| adresseLandKodeAlfa2 | addressCountryCodeAlpha2 | Code | 1 | ISO-betegnelsen for land. | |
| adresseLinie1 | addressLine1 | char(50) | 0-1 | c/o reference. Såfremt landeKodeAlfa2 er forskellig fra 'DK', indgår adresselinien uspecificeret i den udenlandske adresse. | |
| adresseLinie2 | addressLine2 | char(50) | 1 | Gade, husnummer, etage. Såfremt landeKodeAlfa2 er forskellig fra 'DK', indgår adresselinien uspecificeret i den udenlandske adresse. | |
| adresseLinie3 | addressLine3 | char(50) | 0-1 | Såfremt landeKodeAlfa2 er forskellig fra 'DK', indgår adresselinien uspecificeret i den udenlandske adresse. Anvendes ikke hvis landeKodeAlfa2 er 'DK' | |
| adresseLinie4 | addressLine4 | char(50) | 0-1 | Såfremt landeKodeAlfa2 er forskellig fra 'DK', indgår adresselinien uspecificeret i den udenlandske adresse. Anvendes ikke hvis landeKodeAlfa2 er 'DK' | |
| adresseLinie5 | addressLine5 | char(50) | 0-1 | Såfremt landeKodeAlfa2 er forskellig fra 'DK', indgår adresselinien uspecificeret i den udenlandske adresse. Anvendes ikke hvis landeKodeAlfa2 er 'DK' | |
| adressePostboks | addressPostboxReference | Char(4) | 0-1 | Nummer eller anden identifikation jf. Post Danmark | |
| adressePostNr | addressZipCode | char(4) | 1 | Et postnummer består af fire cifre. Refererer til klassen 'PostDistrikt'. Herfra kan 'postDistrikt' hentes (tekstbetegnelse for distriktet) | |
| adresseByNavn | addressCityName | char(34) | 1 | Det fastsatte bynavn præciserer beliggenheden inden for en kommune eller postdistrikt. Et evt. bynavn er en nødvendig del af den fuldstændige og korrekte adresse. Et bynavn skal fastsættes når der findes ens eller enslydende vejnavne i postdistriktet eller kommunen. byNavn kan bestå af indtil 34 tegn. | |

Beskrivelse af elementet ScannedAppendix

| Attribut Dan | Attribut Eng | Type | Kardinalitet | Kommentar og reference til FESD datamo- del | Mapning til NOARK4 |
|------------------------|--------------------------|-----------|--------------|--|-----------------------|
| skannetAppendixFilNavn | scannedAppendixFileName | Char(50) | [1] | Dette element indeholder filnavnet for det tilhørende indskannede dokument. Dokumentet skal ligge i samme filkatalog som FES-DScan.xml | |
| skannetAppendix | scannedAppendixReference | Char(50) | [0..n] | Hvert bilag lagres som selvstændigt Dokument under JournalPosten. Se detaljeret beskrivelse nedenfor. Sammenknytningen af dokumenterne sker ved, at XML-filen udpeger både hoveddokumentet og bilagene ved deres filsystemnavne. | |
| appendixTitel | appendixTitleText | Char(255) | [0..1] | Dette felt benyttes som forslag til Dokument feltet DokumentTitel. | |

Beskrivelse af elementet Rendition

| Attribut Dan | Attribut Eng | Type | Kardinalitet | Kommentar og reference til FESD datamo- del | Mapning til NOARK4 |
|-----------------|-------------------|----------|--------------|--|-----------------------|
| variantFilNavn | renditionFileName | Char(50) | [1] | Dette element indeholder filnavnet for varianten. Dokumentet skal ligge i samme filkatalog som FESDScan.xml | |
| variantType | renditionType | Char(4) | [1] | Varianttypen benyttes til at angive hvilket format den tolkede variant af hoveddokument eller bilag er. Tilladte værdier er TXT eller PDF. | |

Beskrivelse af elementet ScannedDocumentExtension

| Attribut Dan | Attribut Eng | Type | Kardinalitet | Kommentar og reference til FESD datamo- del | Mapning til NOARK4 |
|--------------------------|--------------------------|------|--------------|---|-----------------------|
| skannetDokumentUdvidelse | scannedDocumentExtension | ANY | [0-1] | Elementet indeholder udvidelser til skemaet, med yderligere beskrivende metadata om dokumentet. | |

Beskrivelse af elementet ScannedDocumentContents

| Attribut Dan | Attribut Eng | Type | Kardinalitet | Kommentar og reference til FESD datamo- del | Mapning til NOARK4 |
|------------------------|-------------------------|------|--------------|--|-----------------------|
| skannetDokumentIndhold | scannedDocumentContents | ANY | [0-1] | Elementet indeholder udvidelser til skemaet, dækkende indholdet af indskannede blanket- ter. | |

Bemærk:

Et input til ESDH-systemet bestående af en XML-fil samt ét eller flere dokumenter tages ind i ESDH-systemet, når det afleveres som ovenfor beskrevet. Endemålet med afleveringen til ESDH-systemet er, at der oprettes én JournalPost for hvert hoveddokument, og ét eller flere Dokumenter (en for hvert bilag) hørende til JournalPosten. Der oprettes ikke nødvendigvis en JournalPost direkte ved aflevering til ESDH-systemet, da der kan mangle metadata, som er obligatoriske før en JournalPost kan oprettes. ESDH systemet kan derfor vælge at opbevare de indlæste data i temporære tabeller, indtil nok metadata er registeret til at kunne oprette en JournalPost (Et krævet felt, før en JournalPost kan oprettes, er fx JournalPostSagID, som typisk ikke kan påføres dokumentet, før det er blevet visiteret til journalisering i ESDH-systemet). Hvis det indskannede dokument er opmærket med nok metadata til at opfylde minimumsregistreringen for en JournalPost, kan ESDH-systemet vælge at oprette JournalPosten direkte uden temporær lagring.

Der stilles ingen krav til denne midlertidige opbevaring i ESDH-systemet, som typisk benyttes i fordelingsfasen, indtil et dokument endeligt kan journaliseres på den rigtige sag.

Attributter i FESDScannedDocumentType refererer for en stor del direkte til et felt i JournalPost og kan lagres direkte i de respektive JournalPost felter. Der skal selvfølgelig ske en validering af feltindholdet mod de respektive JournalPost-felters domæne inden lagring. Nedenfor beskrives hvert felt nøjere.

4.3 Afledte krav til kernesystemet

Kravene til skanningsmodulets sammenhæng med kernesystem, herunder den ønskede fleksibilitet med hensyn til tilrettelæggelsen af proceduren, stiller en række supplerende krav til kernesystemet:

- Kernesystemet skal kunne hente de indskannede dokumenter i den ovenfor beskrevne filstruktur
- Sammenknytningen af bilag og hoveddokument sker ved, at bilagene oprettes som selvstændige dokumenter og tilknyttes samme JournalPost som hoveddokumentet.
- Kernesystemet skal kunne benytte felterne RecipientUserReference og RecipientUnitReference til fordeling af dokumentet til sagsbehandler/afdeling, hvis disse er udfyldte
- Kernesystemet skal kunne tilbyde mulighed for optisk kvalitetskontrol, herunder mulighed for at zoome ind og ud af de indskannede dokumenter.

Bemærk også vedørende denne høring:

OIOXML-skeamer udarbejdes efter afsluttet høring, idet høringen kan påvirke udvalget af dataelementer (attributter) og dermed også de tilhørende OIOXML-skemaer

5 Bilag A

Anvendte typer i FESD-modellerne

FESD-modellerne er udarbejdet i UML med det formål at beskrive den logiske informationsarkitektur i en FESD-løsning.

UML-modellen er opbygget af et antal klasser, der igen er opbygget af relationer til andre klasser og primitive typer, så man kan opfatte UML-modellen som opbygget af byggesten, hvoraf den mindste er de primitive datatyper.

I UML-modellen beskriver de primitive datatyper det logiske domæne, som datatypen kan antage, men ikke hvordan den fysiske repræsentation af data skal være.

En gennemgang af de FESD-modeller, der på nuværende tidspunkt foreligger, viser, at vi har anvendt nedenstående primitive datatyper. I skemaerne ses for hver primitiv datatype, der anvendes i FESD-UML-modellen, den tilsvarende datatype for hhv. SQL og XSD.

integer

Benyttes til angivelse af heltal.

Der kan benyttes en angivelse af max længde – hvis intet er angivet, vil domænet være i intervallet mellem -2.147.483.648 og 2.147.483.647. Anvendte længder i den nuværende model: 2, 3, 4, 6, 8, 10.

Benyttes alene til naturlige tal (positive heltal).

Benyttes også til at danne identifikator med udfaldsrum 1 – 2000000000.

Nedenstående skema viser eksempler på heltalstyper og deres repræsentation i hhv. SQL og XSD

| UML Type | Beskrivelse | SQL datatype | XSD datatype |
|--------------------|--|-------------------------|--------------|
| Integer | Heltal i rummet mellem - 2.147.483.648 og 2.147.483.647, begge tal inklusive. Den maksimale længde angives i modellen (*Hvordan ?*) | NUMBER(Maksimal længde) | xsd:int |
| NonNegativeInteger | Naturlige tal Heltal i rummet mellem 0 og 2.147.483.647, begge tal inklusive | NUMBER(Maksimal længde) | xsd:int |

boolean

Er en grundtype i UML.

| UML Type | Beskrivelse | SQL datatype | XSD datatype |
|----------|---|--------------|--------------|
| Boolean | Udfaldsrummet er binært true / false (eller rigtigt / forkert). | BOOLEAN | xsd:boolean |

Identifikationstyper

| UML Type | Beskrivelse | SQL datatype | XSD datatype |
|-----------|---|--------------|---|
| URI | Enhver mulig lovlig URI. Der kan evt. anvendes en max-længde. | STRING | xsd:anyURI |
| HttpURI | Lovlig http(s)-adresse | STRING | xsd:anyURI Evt. restrictions der afgrænser til http. |
| MailToURI | Smtp-adresse på en mailmodtager. | STRING | xsd:anyURI Evt. restrictions der afgrænser til mailadresser. |
| UUID | Identifikator på objekt | UUID | xsd:anyURI Evt. restrictions der afgrænser til UUID. |

char

Benyttes til karakterstreng af varierende længde, men med en defineret maksimal længde. Anvendte længder: 1, 2, 3, 4, 15, 16, 20, 34, 40, 50, 60, 70, 110, 120, 255.

Benyttes også nogle steder til at angive en boolskværdi.

Bør ændres så typen alene bruges til tekststreng:

| UML Type | Beskrivelse | SQL datatype | XSD datatype |
|---------------|--|--------------------------|--|
| Char | Den minimale og maksimale længde er angivet i modellen. | VARCHAR(maksimal længde) | xsd:string Med restriction på den maximale længde |
| ShortText | Anvendes til felter der indeholder en kort beskrivende tekst | VARCHAR(70) | xsd:string Med restriction på den maximale længde |
| Text | Anvendes til en beskrivende tekst med en fast længde. | VARCHAR(255) | xsd:string Med restriction på den maximale længde |
| Code | Anvendes til at beskrive en kode, der er nøgle/fremmednøgle i en opslags-tabel. | VARCHAR(2) | xsd:string Med restriction på den maximale længde |
| Name | Anvendes til felter der indeholder navne, der kan opfattes som brugervendte nøgler. | VARCHAR(70) | xsd:string Med restriction på den maximale længde |
| ReferenceText | Anvendes til at felter der indeholder tekster, der af brugerne opfattes som fremmednøgler. | VARCHAR(40) | xsd:string Med restriction på den maximale længde |

date, dateTime og time.

Typen 'Date' bruges til at angive dato.

Typen 'DateTime' bruges til 'Dato og tidspunkt', også i betydningen 'tidsstempel'.

Typen 'Time' bruges til 'klokkeslæt'.

| UML Type | Beskrivelse | SQL datatype | XSD datatype |
|----------|------------------------------|--------------|--------------|
| Year | Årstal | VARCHAR (4) | xsd:gYear |
| Date | Dato | DATE | xsd:date |
| DateTime | Dato og tidspunkt | DATETIME | xsd:dateTime |
| Time | Tidspunkt uden datoangivelse | TIME | xsd:time |

string

Er en grundlæggende type i UML.

| UML Type | Beskrivelse | SQL datatype | XSD datatype |
|----------|---------------------------------|--------------|--------------|
| String | Tekststreng af vilkårlig længde | STRING | xsd:string |

float

Defineres som sådan:

| UML Type | Beskrivelse | SQL datatype | XSD datatype |
|----------|--|--------------|--------------|
| Float | Decimaltal hvor længde og præcision begrænses til en samlet størrelse på 16-bit. | DECIMAL(X,Y) | xsd:float |

