Remissvar, FormatE, RA-KS 2021-00018, tidigare dnr RA 22-2018-791

Remissvaret är sammanställt av ArkivIT:s arkivarie- och utvecklarkonsulter genom Leif Pettersson, Sam Ekenkrantz, Anastasia Pettersson och Magdalena Sjödahl.

Remissvar, FormatE

ArkivIT grundades år 2010. Bolaget har idag drygt 50 medarbetare, som arbetar inom arkiv, IT och informationssäkerhet och dataskydd. Vårt huvudkontor ligger i Stockholm, vi har ett lokalkontor i Norrköping, och våra konsulter arbetar över hela landet. Våra kunder är främst kommuner, regioner och statliga myndigheter.

Vi är inte en av de utpekade remissinstanserna men önskar ändå framföra våra tankar utifrån den breda erfarenhet våra medarbetare har inom informationshantering.

ArkivIT har inte haft möjlighet att läsa och sätta sig in i samtliga dokument fullständigt. Vi svarar här endast på de delar som vi har haft möjlighet och tid att analysera på ett utförligt sätt. Vi har särskilt valt att fokusera på de föreslagna RA-FS:arna.

Övergripande kommentarer

Vid läsningen av de textdokument som nu remitteras av Riksarkivet är den högt satta ambitionsnivån från myndighetens utredare tydlig. Önskan och ambitionen att på ett gediget och omfattande sätt utreda formatfrågan genomsyrar samtliga dokument.

ArkivIT är mycket positiva till stora delar av utredarens resonemang om öppna format samt att lämpliga format ska användas redan när handlingarna framställs. Vi stödjer att arkiv- och bevarandefrågan inte kan hanteras i efterhand, utan är av yttersta vikt genom hela informationens livscykel för att kunna säkerställa att information finns tillgänglig även på längre sikt. I den digitala världen är livslängden på vissa format betydligt kortare än i den analoga vilket gör att formatfrågan inte bara är ett framtida problem utan en ständig verksamhetsfråga för att säkra tillgång till information som behövs för att säkerställa kärnverksamheter.

I den analoga pappershanteringen har arkivering varit en fråga som i väldigt många fall hanterats i slutet av en ärendeprocess och då ofta av arkivarier och andra experter inom informations-hantering. I och med digitaliseringen behöver fler expertområden engageras i bevarandefrågan i fler och tidigare processkeden. Beslut behöver fattas om format, metadata, konverteringar,

migreringar och strukturer utifrån handlingstyp. ArkivIT stödjer därför tanken på ett mer flexibelt och bredare ramverk än dagens RA-FS.

Omfattningen på remissen i kombination med bland annat en otydlig disposition, egenkonstruerade begrepp och ett komplext ämnesområde gör dock utredningen och regelverket svårläst. Då ett stort antal yrkesgrupper inom offentlig sektor berörs av remissen borde en pedagogisk utformning som tydligt vägleder läsaren genom de delar som är relevanta ha prioriterats liksom en rimlig avgränsning i omfattningen av remissen. Läsanvisningarna som bifogats remissen har säkerligen syftet att underlätta för läsaren men även de blir väldigt omfattande och för otydliga i sin utformning.

I språklagen (2009:600) 11 §, står det att: ”Språket i offentlig verksamhet ska vara vårdat, enkelt och begripligt.” ArkivIT förstår delvis utredarens ambition att tydliggöra och utveckla språkbilden inom området men vårt intryck är att ett tydligare och mer entydigt språk kunde ha använts på många platser för att öka läsbarheten och den breda förståelsen av förslagen och föreskrifterna. Ett exempel är ordet ”fakultativ” som förekommer ett antal gånger i texterna. I stället för fackuttrycket ”fakultativt” bör synonymer användas som ”valfri” eller ”frivillig”.

ArkivIT hade även önskat en större tydlighet och skiljelinje i förslagen och rekommendationerna gällande lagringsformat och tolkningsformat. Ett exempel är stegen mellan UTF-8 till XML, vidare till FGS. Förtydligande gällande gränserna mellan filformat, tolkning och nyttjande hade underlättat läsningen och tillämpningen av bestämmelserna.

Även referenserna till regelverket i WCAG och gällande tillgänglighet bör ha varit tydligare och i högre grad motiverats utifrån syftet med bestämmelserna i det remitterade underlaget.

I konsekvensutredningen resonerar Riksarkivets utredare om tillgång till olika kompetenser och resurser av relevans för digitalisering och hantering av elektroniska handlingar både idag och i en överskådlig framtid. Resonemanget är långt och omfattande, men vår bedömning är att det är av yttersta vikt att en expertmyndighet som Riksarkivet på ett tydligt sätt vägleder andra myndigheter – särskilt som det idag i många organisationer är brist på rätt resurser i den omfattning som behövs för respektive verksamhets informationshantering. I resonemanget saknar vi även kompetensområden, som exempelvis arkivarier. För att främja en effektiv resursfördelning, utifrån befintlig tillgång på olika kompetenser, i hela offentlig sektor bör Riksarkivet arbeta fram tydliga föreskrifter och regelverk för berörda målgrupper med eventuellt tillhörande vägledningar om behov finns.

Arkiv- och bevarandefrågan är en av många frågor som ska hanteras och lösas i samband med val av metoder för framtagande av elektroniska handlingar. Frågan bevakas och följs upp i många organisationer av redan idag hårt belastade resurser. Riksarkivets utredare anser att de nya föreskrifterna inte nödvändigtvis kommer medföra kostnader för andra organisationer än för Riksarkivet. ArkivIT förstår och håller delvis med om resonemanget att verksamheter inte kan skylla eventuella konverteringskostnader på de tillkommande föreskrifter, utan att det lagreglerade kravet på att framställa handlingar i arkivbeständiga format funnits även innan föreskrifterna. ArkivIT anser dock att de föreslagna föreskrifterna behöver ses över igen för att minska risken att de leder till omfattande och kostsamma utredningar hos enskilda organisationer. Sverige är ett förhållandevis litet land där resurser inom den offentliga verksamheten bör användas gemensamt och på ett kostnadseffektivt sätt. Inom detta område bör Riksarkivet tillsammans med bland annat Myndigheten för digital förvaltning, DIGG, arbeta för gemensamma och likriktade ramverk och riktlinjer. ArkivIT:s bedömning är att befintliga förslag inte uppfyller detta.

Ett exempel på där en gemensam lösning bör upprättas är åtkomst och lagring av specifikationer för olika format. För de format som är rekommenderade av Riksarkivet bör detta kunna lösas nationellt och gemensamt för hela offentlig sektor. Detta även för att underlätta kvalitetssäkring och korrekt formatvalidering.

I det remitterade underlaget finns en uttalad ambition om att myndigheter med hjälp av föreskrifterna ska kunna ställa högre och tydligare krav på leverantörer och utvecklare av IT-baserade verksamhetssystem. För att denna ambition ska kunna förverkligas behöver det finnas en nationell samsyn angående *vilka krav* som bör ställas samt hur dessa ska vara *specificerade*. Såsom ArkivIT uppfattar underlaget från Riksarkivet ska varje arkivmyndighet, eller annan funktion med likartad funktion, samt varje enskild myndighet tolka och själv tillämpa de nu föreslagna föreskrifterna för att därefter sammanställa policy, riktlinjer och specifikationer som ska gälla för den enskilda verksamheten. ArkivIT uppfattar dock att de tolkningsutrymmen som finns i regelverket, framför allt i RA-FS 2022:X bilaga 3, kommer att innebära ett för stort spann i tillämpningarna. Så som föreskrifterna är formulerade och utformade i det remitterade förslaget anser ArkivIT att de inte kommer att främja en gemensam och enhetlig kravbild för liknande system inom offentlig sektor. I stället kommer den situation som finns idag att fortsätta med över 300 myndigheter som ställer sina separata krav beroende på hur de har tolkat Riksarkivets texter.

Vår uppfattning är även att ytterst få befintliga system idag, inklusive de dokument- och ärendehanteringssystem som förekommer på svenska marknaden, kan uppfylla samtliga av de krav som ställas i de föreslagna föreskrifterna för allmänna handlingar. Det kommer att krävas stora insatser för leverantörerna att skapa helt ny funktionalitet i sina system och om det då inte finns en samlad och överenskommen bild *hur* de nya funktionaliteterna ska fungera och *vad* de förväntas göra kommer den situation som finns idag inte att förändras till det bättre.

ArkivIT saknar även resonemang, för respektive föreslaget format, om klimat, energi och miljö. Energi- och klimatfrågan tillsammans med ekonomiska analyser är i många fall en av de aspekter som vägs in när lagrings- och formatfrågor diskuteras internationellt idag för omfattande data- och informationsmängder. Olika intressen gällande till exempel komprimering ställs i många fall mot varandra. Det finns flera initiativ internationellt som har fokus på just digitalt bevarande och miljöhänsyn. Bland annat kan nämnas projektet Archiver som arbetar med hantering av forskningsdata och inom det arbetet tittat på hållbarhetsfrågan[1] samt debattartikeln ”Toward Environmentally Sustainable Digital Preservation” i The American Archivist 2019. Debattartikeln berör bland annat energiförbrukning men också andra miljöpåverkande delar i samband med digitalt bevarande.[2] Inom EU har det inletts ett förarbete för att ta fram riktlinjer som planeras styra upphandling av datacentraler, serverhallar och molnlösningar (se vidare ”EU green public procurement criteria for data centres, server rooms and cloud services”).[3]

ArkivIT saknar även tydlig analys av kryptering och lösningar av detta vid lagring och för enskilda filer. Krypteringsproblematiken nämns och tillåts i samband med transport av filer och kommunikation, samt vid e-underskrifter. Frågan är dock ständigt aktuell även vid lagring och hantering av elektroniska allmänna handlingar.

Som utredaren själv för ett visst resonemang om i det utsända underlaget tillkommer ständigt nya filformat, även om utredaren vänder sig mot uttryck som att ”tekniken går för fort”. ArkivIT ser en utmaning i, för Riksarkivet, att underhålla och förmedla förändringar och tillkommande (och frånfallande) rekommendationer för format. Ett exempel som ArkivIT identifierat är Wiki. Det saknas idag en överenskommen standard för Wiki Markup men Wiki används idag flitigt för till

[1] Projektet hade ett webbinarium i november 2021 med rubriken ”Environmentally sustainable digital preservation – moving from theory to practice”, https://www.archiver-project.eu/environmentally-sustainable-digital-preservation-moving-theory-practice [2022-01-27].
[2] https://dash.harvard.edu/handle/1/40741399 [2022-01-27].
[3] https://ec.europa.eu/environment/gpp/pdf/20032020_EU_GPP_criteria_for_data_centres_server_rooms_and%20cloud_services_SWD_(2020)_55_final.pdf [2022-01-27].

exempel olika typer av dokumentation. Wiki är ett exempel som skapar frågor som ligger nära till hands det som berörs i Riksarkivets underlag. Är Wiki ett lagringsformat? När och hur används det, finns behov av Wiki? Hur ska det hanteras framöver, till exempel om en Wiki-standard tas fram?

ArkivIT:s önskan och förhoppning är att Riksarkivet hittar tydliga och pedagogiska kanaler för att kommunicera och vägleda övrig offentlig sektor vid förändringar i riktlinjerna. Riksarkivets utredare skriver angående frågan om att förvalta det remitterade underlaget och föreskrifterna att

> ”Den fortsatta uppdateringen är relativt enkel så länge det görs regelbundet. Av betydelse är därför att underlaget förvaltas, och att det finns resurser och rutiner för den långsiktiga förvaltningen. Det finns sedan förslag på hur på hur Riksarkivet kan effektivera förvaltningen. Till exempel, dels strukturera informationen, dels systematisera och automatisera vissa arbetsprocesser, dels inrätta arbetsgrupper som ansvarar för varje kategori av form och funktion.”.

ArkivIT tror att denna aspekt inte ska underskattas av Riksarkivet och att en god förvaltningsorganisation är avgörande för att föreskrifterna ska kunna efterlevas och vara relevanta i framtiden.

Riksarkivets föreskrifter och allmänna råd om arkivrättsliga krav vid framställningen av allmänna handlingar, ArK

I detta avsnitt har ArkivIT samlat synpunkter, tankar och resonemang om *Riksarkivets föreskrifter och allmänna råd om arkivrättsliga krav vid framställningen av allmänna handlingar, ArK*.

- Det bör framgå redan i dokumentets titel att föreskrifter och allmänna råd som hanteras i detta dokument endast rör information i elektroniskt format. Detta framgår först i förslagets första stycke – ”… föreskrifter om arkivrättsliga krav på användningen av materiel och metoder vid framställningen av elektroniska handlingar (upptagningar för automatiserad behandling)…”.[4]

[4] RA-FS 2021:Y, sidan 1.

- Vi önskar att den text som anges i avsnittet ”Allmänna råd” om synen på allmänna handlingar och dess framställan på sidan 2 i RA-FS 2021:Y förtydligas. Vi ser exempelvis inte någon skillnad mellan att spara en video i flera olika format eller att spara ett DocX-dokument till PDF/A. Främst är det texten ”…DocX framställs som PDF/A…” som kan anses förvirrande. Texten kan tolkas som att en användare i ett DocX-dokument väljer att skriva ut detta som en PDF/A-fil vilket inte skiljer sig från att spara en videofil i olika filformat. Det bör även hänvisas till gällande lagstiftning om bevarande av allmän handling, i stället för att föra ett resonemang om detta i denna typ av föreskrift.

- Vi anser det högst problematiskt att Riksarkivet väljer att inte följa Tryckfrihetsförordningens definition av allmän handling. Det bör framgå att definitionen i dokumentet rör ”Allmän handling i elektroniskt format”.[5] I föreskriften uppfattar även ArkivIT att begreppet *elektronisk handling* används synonymt med vad som definierats som allmän handling i dokumentets andra kapitel.

Bilagor till RA-FS 2021:Y - BILAGA 1 ARKIVRÄTTSLIGA KRAV FÖR OLIKA FALL

I detta avsnitt har ArkivIT samlat synpunkter, tankar och resonemang om *Bilagor till RA-FS 2021:Y - BILAGA 1 ARKIVRÄTTSLIGA KRAV FÖR OLIKA FALL.*

- **Avsnitt 1.1.1 Allmänna handlingar med bilder.**
  ArkivIT saknar kostnads- och genomförandeanalys för de föreslagna riktlinjer för bildhantering och skanningsrutiner som föreslås i föreskrifterna.

  I avsnitt 1.1.1. Allmänna handlingar med bilder framgår det, med en hänvisning till RA-FS 2021:X, att handlingar med bilder ska framställas i format utan komprimering alternativt med förlustfri komprimering:

  > ”Ett tekniskt skick ska uppfylla vad som sägs i fall #1 om det avser en handling som ska bevaras för all framtid och, eller överlämnas till en arkivmyndighet, eller annan med motsvarande ansvar. Med andra ord, förlustgivande komprimering är endast tillåtet för fall

[5] Ibid.

> handlingar inte blir allmänna eller blir allmänna men som får gallras med stöd av lag."

Ett exempel här är fotografier i elektroniskt format. En stor andel av de fotografier som tas idag blir, utan att användaren kan påverka det, automatiskt JPEG-filer då de vanligaste fotografiska applikationerna som används inom verksamheterna är smarta mobiler. Det kan handla om verksamhetskritiska bilder med stort värde ur ett rättskipningsperspektiv och med ett stort historiskt värde.

ArkivIT ser en otydlighet i förslaget huruvida denna typ av bilder inte ska tillåtas och hur en framtida eventuell löpande konverteringsprocess till TIFF eller PNG skulle se ut och bli försvarbar ur ett kostnads- och hållbarhetsperspektiv. Vi saknar även en analys om huruvida det är hållbart att väsentligt öka filstorleken även om kostnaderna för datalagring historiskt har sjunkit.

I samma avsnitt finns även följande krav för skannade dokument:

> "För fall bilder representerar texter *ska* [kursiverat av ArkivIT] dessa bilder göras till maskinläsbar text. Till exempel, skannade kontorsdokument."

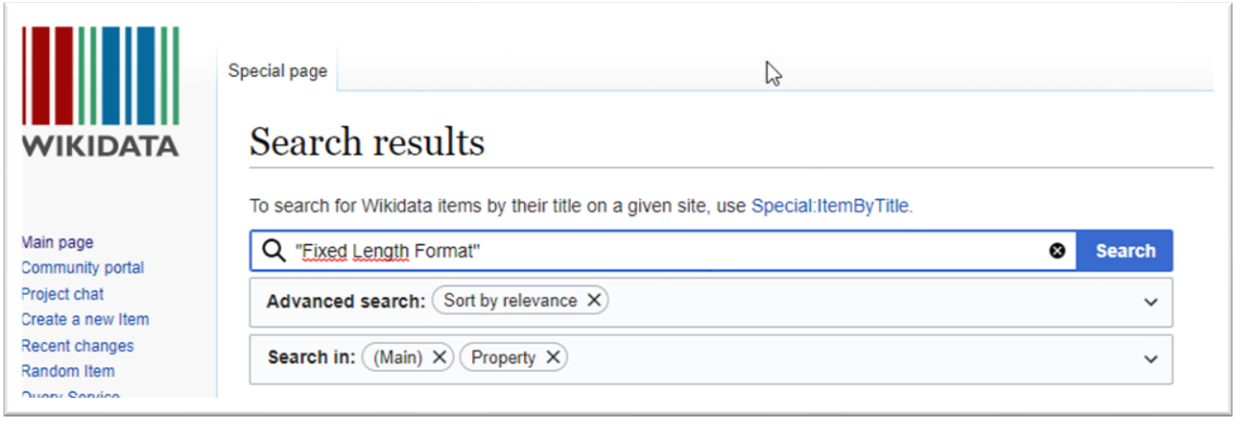
Vår erfarenhet är att en absolut majoritet av de myndigheter vi har haft eller har kontakt med *inte* teckentolkar sina skannade dokument idag. Detta kan bero på många anledningar, såsom resurser, kompetens och behov. Det är maskinellt kraft- och energikrävande att teckentolka handlingar. Dagens teknik för OCR-tolkning är också i vissa fall av mycket hög kvalitet men få, om någon, lösningar uppnår en komplett och helt korrekt tolkning av alla typer av texter. Om detta ska vara ett ska-krav bör Riksarkivet därför även bestämma ett kvalitetskrav för teckentolkningen. Vad skulle vara ett rimligt krav? Mer än 95 % korrekt teckentolkning per sida? Om 95 % uppnås, vilket inte är enkelt, för en normal A4-sida med cirka 4 500 tecken skulle det betyda att omkring 225 tecken per sida skulle bli feltolkade, vilket i sin tur kommer att påverka sökbarheten. Detta skulle också få följder för vilka krav som behöver ställas på skanningskvaliteten.

- **1.1.2. Allmänna handlingar med databaser eller datauppsättningar**

  I avsnitt 1.1.2. Allmänna handlingar med databaser eller datauppsättningar tas formatfrågan upp för bland annat register och andra förteckningar. Här anges att det tekniska skicket *ska* vara i FLF alternativt DSV. Dessa två filformat anser ArkivIT inte vara realistiska alternativ för en majoritet av myndigheterna som framställer handlingstyperna register och förteckningar.

  FLF (Fixed Length Format) kan ses som ett utdöende filformat som ytterst få applikationer idag kan framställa. Formatet används av ett fåtal specialiserade applikationer. En sökning i PRONOM ges följande resultat.

En sökning i Wikidata ger följande resultat.



I en av ArkivIT ackumulerad lista med filändelser som har hittas vid ett antal analyser av olika filservrar och som innehåller analyser av 10-tals miljoner filer finns idag 5 954 unika filändelser, *.FLF förekommer inte en gång i den listan.

Angående DSV (Delimiter Separated Values) går det idag inte att spara i det formatet i varken Microsoft Excel eller till exempel LibreOffice Calc. Det blir sedan förvirrande för oss som läsare när vi jämför de texter som finns angående CSV i RA-FS 2021:X. På sidan 60 i dokumentet ges följande information i avsnittet 3.3.2 Databaser och datauppsättningar:

> 1. Filformat | CSV | Får | ”CSV är en varians av DSV, och är generellt lämpligt för att representera tabulär data i textformat.”

I posten under, nummer 2, ges förklaringen att DSV är ett generiskt format som bland annat har varianserna CSV och TSV.

Ytterligare förvirrande blir det för läsaren när det längre ned i RA-FS 2021:Y i avsnitt ”1.1.5. Allmänna handlingar med kalkylblad” anges att för kalkylblad som endast är en sammanställning av tabulär data utan formler, något som skulle kunna tolkas som ett register eller en förteckning, ska det tekniska skicket vara DSV men får *inte vara* i FLF.

I RA-FS 2021:X, sidan 80, anges det ”Får (villkorat)” för filformaten CSV och TSC med nästan exakt samma förklarande text – ”För fall kalkylblad representerar tabulär data, utan formler eller andra samband som har väsentlig betydelse för innehållet.”

Vad är det som gäller här? ArkivIT anser att krav som ställs av en myndighet ska vara realistiska och med rimliga insatser vara genomförbara. För att kunna ställa realistiska ska-krav i det här fallet behöver dessa förhålla sig till de applikationer och programvaror som generellt används av myndigheter och de filformat som dessa kan hantera och/eller producera.

- **1.1.3 Allmänna handlingar med datafiler**

  I avsnittet 1.1.3 Allmänna handlingar med datafiler tas frågan om filnamn upp. Här formuleras bland annat regeln att:

  > ” I filnamnet får punkt endast förekomma som skiljetecken för att signalera avgränsningen till filändelsen.”.

  Som exempel ges sedan bland annat ”dokument_docx.a1a.pdf”.

Generellt har ArkivIT invändningar till hur kraven för filnamn är satta. Vi anser att det är mycket viktigt med filnamnsstrategier för *hur* filnamn ska vara uppbyggda. Att ändra

filnamn i efterhand kräver först en konsekvensanalys. Det ursprungliga filnamnet är i vissa fall en typ av metadata alternativt en del av handlingen där det framgår till exempel versionsangivelser, rubriker och datum. Om filnamn ska ändras är det av stor vikt att det tas hänsyn till framtida användare av filerna och till konsekvenserna av den gallring som byte av filnamn kan innebära. Om filnamn är intetsägande och inte rimligt förklarande blir det mycket svårt att använda dessa till exempel efter ett uttag ur ett e-arkiv. Filnamn bör även möjliggöra sortering i exempelvis Utforskaren samt ge en möjlighet att förstå vad filen hanterar innehållsmässigt. Det regelverk och de exempel som anges i avsnittet är gränsande till omöjliga att tekniskt och automatiserat införa. Ska ett diariesystem behöva integreras med Droid eller JHove för att extrahera data som i sin tur sedan ska struktureras enligt exemplen? Exempel som ”film_h264_aac.mp4” eller ”anteckningar_utf8.txt” är ur ett användarperspektiv mycket olämpliga.

- **1.1.5. Allmänna handlingar med kalkylblad**

  Hanteringen av formler som föreslås i avsnitt 1.1.5. Allmänna handlingar med kalkylblad är enligt ArkivIT:s erfarenhet inte en realistisk hantering. Det är inte ovanligt att formler hämtar data från olika flikar i ett kalkylblad. I RA-FS 2021:X framgår inte med tydlighet hur ODS- eller XLSX-filer faktiskt ska hanteras *om* de kan ses som komplexa, till exempel med formler. Att i klartext skriva ut en formel. Eller att i klartext hantera formler i en separat handling med relation till en ODS/XLSX. Eller att extrahera XML från de XML-filer som ODS/XLSX faktiskt består av för att sedan bädda in dessa extraherade XML-filer i en PDF/A-3. Att utföra detta för ODS/XLSX som skapas i framtiden eller på de filer som redan existerar går inte.

- **1.1.9. Allmänna handlingar med underskrifter**

  ArkivIT har svårt att tolka innebörden och konsekvenserna av avsnittet 1.1.9. Allmänna handlingar med underskrifter. Hur ska exempelvis begreppet ”digital signatur” ses gentemot det i eIdas-förordningen[6] definierade begreppet elektronisk underskrift?

  I avsnittet står det att ”För fall underskriften inte ska gallras, ska bevaras för all framtid,

[6] Europaparlamentets och rådets förordning (EU) nr 910/2014 om elektronisk identifiering och betrodda tjänster för elektroniska transaktioner på den inre marknaden och om upphävande av direktiv 1999/93/EG, vanligen kallad eIdas-förordningen.

och eventuellt överlämnas till arkivmyndigheten, ska i ett och samma tekniska skick omslutas alla format som använts för att åtminstone en gång ha tekniskt kontrollerat underskriften (validering)." Det redovisas sedan en lista med faktorer som är av särskild betydelse. ArkivIT undrar hur detta bör hanteras rent praktiskt. Är det så att denna information ska kunna redovisas också i metadata för handlingar som till exempel har ingått i ett ärende? Om så är förslaget från Riksarkivets sida är vår bedömning och förhoppning att befintliga FGS:er bör ses över för att möjliggöra och underlätta så att metadata, på ett standardiserat sätt, visar att detta har skett och att till exempel validering har gjorts samt när.

- **1.2. Särskilda arkivrättsliga krav/ 1.2.1.1. Offentliggörande av års- och koncernredovisningar**

  I avsnitt 1.2. Särskilda arkivrättsliga krav/ 1.2.1.1. Offentliggörande av års- och koncernredovisningar ställs krav på iXBRL samt XBRL för handlingstyper som års- och koncernredovisningar, revisionsberättelser, med fler ekonomiinriktade handlingar. Detta på grund av att detta krävs med hänvisning till två EU-regelverk – förordningen (EU) 2018/815, bilaga IV samt direktiv 2004/109/EG.

  När ArkivIT läser de dokument som det hänvisas till framgår det att dessa reglerar "…insynskraven angående upplysningar om emittenter vars värdepapper är upptagna till handel på en reglerad marknad…". Det vill säga *börsnoterade bolag*. Vidare ska filformatet vara XHTML med filer uppmärkta enligt IFRS-standard som till exempel XBRL.

  ArkivIT kan inte se att det finns några som helst krav på att *myndigheter* är förordnade att följa de regelverk som är uppsatta av EU. ArkivIT önskar därför förtydligande från Riksarkivet om dessa krav ska gälla även för myndigheter, samt eventuellt en omformulering av kraven angående filformat för den här typen av handlingar.

Riksarkivets föreskrifter och allmänna råd om tekniska krav vid framställning av elektroniska handlingar, TeK

I detta avsnitt har ArkivIT samlat synpunkter, tankar och resonemang om *Riksarkivets föreskrifter och allmänna råd om tekniska krav vid framställning av elektroniska handlingar, TeK.*

- Den lista som redovisas i bilaga 1 av olika standarder (och icke-standarder) är värdefull. Den kommer att bli en mycket bra referenssamling för det fortsatta arbetet med att hantera filformat för digitalt bevarande.

- Även påföljande bilaga 2 är en värdefull förteckning av de olika verktyg om finns för verifiering och validering.

Bilaga 3

- ArkivIT önskar förtydligande gällande och att punkterna, sidan 53, i bilden nedan slås samman. Det känns förvirrande att samma filformat (version av PDF) förekommer i båda posterna med en nyans i den förklarande texten. Till exempel, varför anges PDF/X-3:2003 i post nummer 1 och endast PDF/X-3 i post 2. Vad är skillnaden? Finns det fler versioner av PDF/X-3?

3.3.1.3. *Förtryck och tryckteknik*

Tabell 15 Särskilda tekniska krav på formen och funktionen bild för speciella fall som förtryck och tryckteknik.

#	Teknisk egenskap Teknisk process	Specifikation Implementering	Teknisk kontroll	Krav	Skäl
1.	Filformat	PDF/X-1a PDF/X-3:2003		**Får (villkorat)**	En PDF kan överstämma med kraven i både delmängd "1a" eller "3:2003" av PDF/X och PDF/A-1 under förutsättning att inga funktionaliteter implementeras från PDF/X som är uttryckligen förbjudna i PDF/A-1.
2.	Filformat	PDF/X-1+1a PDF/X-1a PDF/X-3 PDF/X-4p		**Får (villkorat)**	En PDF kan överstämma med kraven i både delmängd "1", "1a", "3", eller "3p" av PDF/X och PDF/A-2 eller -3 under förutsättning att inga funktionaliteter implementeras från PDF/X som är uttryckligen förbjudna i PDF/A-2 eller -3.

- 3.2.2.4 Tillgänglighet
  Vi saknar tydlighet vad gäller om ett angivet filformat anses vara lämpligt som bevarandeformat eller inte. I det här fallet tas till exempel PDF/UA upp med kravet "Bör". Betyder det att PDF/UA är godkänt som ett bevarandeformat? Senare i bilagan i avsnittet "3.3.6.2. Geografi" omnämns PDF/UA på nytt men då helt lösryckt i samband med PDF/A-4.

> ” PDF/A-4 skiljer sig väsentligt från tidigare versioner av PDF/A. Till exempel, nivåerna ”a”, ”b”, ”u” finns inte längre, medan Javascript är numera tillåtet, och PDF/UA är avsedd att uppfylla behoven av och kraven på logisk och semantisk struktur.”

Här är kravet ”Bör inte”. ArkivIT önskar förtydligande om vad PDF/UA har för koppling till PDF/A-4 i samband med ”geografi”.

Exakt samma inledning återkommer på nytt under avsnittet ” 3.3.6.3. Kontorsdokument” och också då i samband med PDF/A-4. Då med kravet ”Bör inte (villkorat). Tyvärr omnämns inte PDF/UA som en egen entitet i bilagan vilket leder till en osäkerhet.

- Ett annat exempel på otydlighet är när filformat omnämns på ett antal olika platser i bilagan. Ett exempel är MPEG-4 AVC som omnämns i 12 olika sammanhang. Formatet förekommer i delarna *Fall*, *Skäl* och *Exempel*.
    - 3.2.3. Komprimering, punkt 1, anges under Exempel, Krav ”Får (villkorat).
    - Samma avsnitt, punkt 4, anges under Exempel, Krav ”Får (villkorat).
    - Samma avsnitt, punkt 6, anges i ett exempel, Krav ”Får inte”.
    - 3.2.4. Licenser och patent, punkt 1, anges i exempel, Krav ”Bör inte”.
    - Samma avsnitt, punkt 2, anges som exempel, Krav ”Bör inte”.
    - Samma avsnitt, punkt 3, anges som exempel, Krav ”Får (villkorat)”.
    - 3.3.6.4. Publicering, punkt 2 (här saknas punkt 1), anges i Skäl för formatet EPUB som rekommenderat format, krav ”Får (villkorat).
    - 3.3.7.1. Generella fall, punkt 7, anges i Skäl för formatet MP4, Krav ”Bör inte”.
    - Samma avsnitt, punkt 15, anges i Skäl för filformatet AV1 där det anges att AV1 bör kunna ersätta MPEG-4 AVC.
    - Samma avsnitt, punkt 18, anges under Specifikation/Implementering, krav ”Bör inte”.
    - Samma avsnitt, punkt 19, anges i Skäl för MPEG-4 Visual, Krav ”Bör inte”.
    - Samma avsnitt, punkt 22, MPEG-4 AVC anges åter i Specifikation/Implementation, Krav ”Bör inte”.

För att underlätta för de myndigheter som förväntas använda den här bilagan i sitt arbete med bevarande av elektroniska handlingar bör en annan struktur övervägas där information kring Specifikation/Implementering redovisas mer samlat.

- I avsnittet ”3.2.3. Komprimering” ges ett exempel i punkt 6. ”Att ladda upp en video i MPEG-4 AVC till en videotjänst på Internet som sedan kodar om den igen med förlustgivande komprimering.” Ska detta tolkas som att *det är förbjudet att använda vissa vanligt förkommande filformat på till exempel sociala medier* som JPEG, MPEG, MP3 med fler komprimerade format (se punkt 4)?

  Vad händer i nästa skede när webbsidor och sociala medier arkiveras och de format som förekommer i dessa ska arkiveras? Enligt den text som anges som Skäl för filformatet WARC (sidan 86) anges följande – ” Hämtade format omsluts i filformatet. Ett tekniskt skick utgörs av alla omslutna format, varför dels *alla eventuella format måste även uppfylla kraven i denna författning* [vår kursivering], dels en implementering inte får exekvera format, antingen inte alls, eller åtminstone inte innan en teknisk kontroll har bekräftat deras format.”.

- Angående filformatet WARC finns det en motsägelse i handlingarna. I RA-FS 2022:X anges på sidan 86 att filformatet WARC får användas villkorat. I RA-FS 2022:Y på sidan 11 anges följande – ”Webbsidor får inte organiseras och sammanställas i WARC.” ArkivIT önskar här ett mer utvecklande resonemang och motivering till denna motsägelse.

- Vad avgör om en XML-fil är ett Informationsformat (EBUCore sidan 73), Metadata (MIX sidan 53) eller ett Dokumentformat (XBRL sidan 66 och RAFGS3 sidan 69)?

  Vad vi kan finna finns inte de tekniska egenskaperna *Informationsformat*, *Metadata* eller *Dokumentformat* definierade i dokumentet.

- ArkivIT ser en stor utmaning för de myndigheter som förväntas tolka och arbeta utefter de riktlinjer som återfinns i avsnittet 3.2.4. Licenser och patent. ArkivIT efterlyser en

beskrivning av hur Riksarkivet förväntar sig att respektive myndighet ska kunna sätta sig in i samtliga applikationers licensmodeller för att sedan avgöra om de kan använda ett specifikt filformat som har framställts av just den applikationen.

ArkivIT tolkar bestämmelserna som att de innebär att filformat kan användas om det kommer från applikation A men inte om det kommer från applikation B. Som referensexempel har en normalstor kommun mellan 200–400 olika system i sina verksamheter som i högre eller lägre grad framställer information som kan ses som allmän och i många fall sedan ska bevaras. Om Riksarkivet har analyserat system med utgångspunkten licenser bör detta publiceras som stöd i myndigheternas arbete.

- 3.3.3.6 Vad innebär rubriken ”Offentlig verksamhet”? Vi förstår inte syftet med avsnittet. Betyder det att de FGS.er som anges är frivilliga när *Krav* är ”Får”? Att myndigheterna vid leverans till Riksarkivet inte behöver följa de FGS:er som är framtagna?

- En del hänvisningar inom bilaga 3 fungerar inte, vilket väsentligt försvårar tolkningen och läsningen av dokumentet. Några exempel är:
    - 3.2.2.2. Omsluta och innesluta, punkt 2. I Exempel finns en hänvisning rörande PDF/A-3 ” Ett kontorsdokument som framställts som PDF/A-3a kan även bifogas det resulterande PDF/A-3-dokumentet, eftersom det finns ett särskilt krav som tillåter en sådan Organisering och sammanställning (s. 78).”. På sidan 78 eller 79 finns sedan inget omnämnande angående PDF/A-3.
    - 3.3.3.5. Kontorsdokument, punkt 1. I Skäl anges i text ” En arkivmyndighet eller annan med motsvarande ansvar bör därför närmare avgränsa och specificera element och attribut för särskilda typer av elektroniska handlingar. För Organisering och sammanställning av webbsidor se det avsnittet (s. 78).”. På sidan 78 eller 79 finns sedan inget omnämnde hur webbsidor ska organiseras eller sammanställas. Samma gäller även punkterna 2 och 3 i avsnittet med samma hänvisning.
    - 3.3.6.6. Webbsidor, punkt 1. I Skäl finns en hänvisning ” För HTML som Informationsformat se det avsnittet (s. 63).” På sidorna 63–65 finns sedan inte HTML upptaget som informationsformat i Specifikation/Implementering.

- ArkivIT önskar förtydligande om hur texterna om ODD och OOXML ska tolkas. Formaten anges både som Dokumentformat (sidan 69) och Filformat (ODD sidan 80 och OOXML sidan 81). Åter hänvisas användaren i *Skäl* för båda filformaten till en sida, ”… som Informationsformat se det avsnittet (s. 63)” där dessa sedan inte representeras.

  På samma sätt hänvisas det som dokumentformat till ” För Organisering och sammanställning av kontorsdokument se det avsnittet (s. 78).” där inget av formaten finns omnämnda. Är det en hänvisning till ZIP (punkt 6)?

  I *Skäl* för den tekniska egenskapen dokumentformat gällande ODD/OOXML omnämns dessa som ”informationsformat”. Informationsformat är en av de benämningar som används som Teknisk egenskap/Teknisk process. Vad gör att de anges som dokumentformat för att sedan benämnas informationsformat när båda är tekniska egenskaper?

- På flera ställen i det remitterade underlaget finns dubblering av texter, det vill säga att samma text/stycke förekommer på olika platser och i olika sammanhang. Ett exempel rör informationsformatet EBUCore. På sidan 73 under avsnittet 3.3.5. Ljud samt 3.3.5.1. Generella fall där följande text återges i *Skäl*:

  > ”EBUCore utökar ”Dublin Core” för att omfatta framställning, hantering, och bevarande av audiovisuellt material. Den är avsedd att vara interoperabel med andra Dublin Core specifikationer, som Europeana, och kan kopplas om till METS. EBUCore har både ett XML-schema och en fullständig RDF-ontologi. Specifikationen är under kontinuerlig utveckling. Det vill säga, att dokumentet är ett ”levande dokument”.”

  Exakt samma text anges därefter för EBUCore på sidan 90 i avsnittet 3.3.7. Rörlig bild och video/ 3.3.7.1. Generella fall. ArkivIT upplever att denna typ av dubblering skapar en otydlighet i dokumentet och hade önskat att det i stället fanns en post för EBUCore där både ljud och video finns representerade i en förklarande text.

- ArkivIT uppfattar att det i dokumentet förekommer motstridiga krav gällande HTML++. I bilagan återfinns följande två olika krav:
    - 3.3.3.5. Kontorsdokument, punkt 2. Här klassas HTML++ som Dokumentformat med kravet ”Får”
    - 3.3.6.6. Webbsidor, punkt 3. Här klassas HTML++ som Filformat med krav ”Bör inte” med följande text i Skäl ” För annan strukturerad information bör övervägas HTML 4.01 Strict eller XHTML Strict. Till exempel, e-post, dokument, texter.”

Författningskommentarer

I detta avsnitt har ArkivIT samlat synpunkter, tankar och resonemang om *Författningskommentarerna.*

- På sidan 201 finns följande text:

  ”Det minimala kravet för att analogvis tillämpa samma förutsättningar för flera verksamheter bör lämpligen vara att det rör sig om åtminstone samma förutsättningar och omständigheter. Till exempel, att exakt samma program och tekniska miljö används för att framställa eller kontrollera samma typer av elektroniska handlingar. Anledningen är att även små variationer i program och den tekniska miljön kan få stora konsekvenser vid framställningen och återgivningen av elektroniska handlingar.”

  Tyvärr är realiteten att det är gränsande till omöjligt att säkerställa exakt samma tekniska miljö. Även om två identiska datorer används av samma fabrikat, samma modell och upphandlade samtidigt kan delar som till exempel moderkort vara från två olika tillverkare vilket gör att de kan fungera lite olika. Om sådana faktorer är av avgörande vikt så behöver vi förhålla oss till detta.

- På sidan 209 talas det om ”tillfälliga mellanled”:

  ”Till exempel, även om en elektronisk handling endast är avsedd att användas och hanteras internt kan det alltid uppstå ett behov att utbyta den med annan verksamhet eller allmänheten. En enkel tumregel är att ju större risk att ändamålen kan förändras under användningen och hanteringen av handlingen eller att handlingen

> kan bli allmän i och med användningen och hanteringen desto mer bör en verksamhet förhålla sig till denna författning."

Vi upplever det här stycket som en stor utmaning inom vissa verksamhetsområden. Ett sådant exempel är entreprenad- och byggprojekt där en stor mängd mycket specifika programvaror används främst under förprojektering, projektering samt genomförande. Andra områden vi kan se där detta kommer vara problematiskt är inom sjukvården och i högskolornas forskningsverksamhet.

- På sidan 210 talas det i ett exempel om att PDF-filer i stort, med rätt handhavande, skulle kunna jämställas med en PDF/A-fil och att detta skulle bekräftas med en teknisk kontroll. Tyvärr finns inte den möjligheten idag. Det verktyg som det hänvisas till i remissen är VeraPDF. VeraPDF kan endast kontrollera filer som är deklarerade som PDF/A och PDF/UA. I teorin skulle en kunnig person kunna ändra detta genom att skapa ett eget skript i VeraPDF som bortser från att deklarationer saknas. Här vore det en klar fördel om det skriptet kunde skapas av en central organisation, såsom Riksarkivet, som sedan kunde spridas till andra.

- På sidorna 1041 och 1042 framställs en bild av att det inte kommer att medföra några ökade kostnader att spara filer i okomprimerade format eftersom kostnaden för minne minskar över tid. Detta är enligt texten också viktigt för att säkra framtida användandet.

  > "Att inte tillämpa förlustgivande komprimering blir därför en nödvändig förutsättning för elektroniska handlingar som ska bevaras för all framtid, eftersom möjliga användningar och hanteringar inte kan förutsägas för en sådan tidsperiod."

  Här hade det varit önskvärt att frågan problematiserades ytterligare genom att väga in bland annat miljöaspekter kring att i vissa fall väsentligt öka minnesanvändandet. Vilka alternativ kan Riksarkivet se här? - mekaniska hårddiskar, band, andra typer av minne som kan komma i framtiden som syntetisk DNA?

ArkivIT ser också en motstridighet i det resonemang som förs längre fram i dokumentet om de utmaningar som skanning innebär. Riksarkivets utredare skriver här om upplösning, färggdjup, ecetera. På sidan 1062 lyfts argumentet fram för att inte

genomföra skanning med hög upplösning (det är en ganska stor skillnad mellan 300 och 600 SPI i filstorlek). Argumentet för en ”lagom” upplösning är att en högre sådan skulle *öka lagringskostnaderna*. Något som det tidigare i texten argumenterats för inte skulle uppstå trots mer omfattande behov av datalagring. Om lagringskostnaderna inte skulle öka, enligt tidigare uttalande, skulle det optimala vara att alltid skanna i 600 SPI och fullfärg, eller åtminstone med 256 gråtoner, eftersom detta helt klart ger en bättre kvalitet och till exempel skulle ta höjd för små teckenstorlekar samt sannolikt förbättra teckentolkning av text (OCR).

Konsekvensutredning

Angående konsekvensutredningen är det mycket bra att frågan lyfts angående att få kontroll över när till exempel filformat ska användas. Att den kontrollen kommer redan när handlingarna skapas. ArkivIT anser att det är en framgångsfaktor för att i ett senare skede kunna spara information över tid och i förlängningen bevara delar av den. Det skulle verksamheterna vinna på i längden.

Däremot håller vi inte med om att det arbetet inte skulle öka kostnaderna, åtminstone i ett inledande skede. Vi tror att de kostnadsmässiga konsekvenserna skulle kunna bli omfattande för myndigheter som tidigare har haft sämre kontroll och styrning i dessa frågor.