

Godkända format för leverans av digitala filer till Landstingsarkivet

Dokumentägare: Utveckling och stöd

Dokumentnamn
SLL LA Formatspecifikation_1_0.docx

Datum
2015-01-28

Versionsnummer
1.0

Innehållsförteckning

Dokumenthistorik	3
1 Inledning	4
2 Godkända leveransformat	4
3 Textbaserade dokument – PDF/A	5
3.1 PDF/A i korthet	5
3.2 Konvertera till PDF/A	6
3.3 Transparens.....	6
3.4 Färgkodning	7
3.5 Hyperlänkar	7
3.6 Formulär.....	7
3.7 Tillgänglighetsanpassade PDF	7
3.8 Skannade handlingar	7
3.9 Validering	8
4 Textbaserade format	8
4.1 XML.....	8
4.2 TEXT	8
5 Bildformat	8
5.1 TIFF.....	9
5.1.1 JPEG-filer.....	10
6 Ljudformat	10
6.1 Broadcast WAVE Format	11
6.1.1 MP3	11
7 Videoformat.....	12
7.1 MPEG-4	12
7.2 MPEG-1	12
7.3 MPEG-2.....	13

Dokumentnamn
SLL LA Formatspecifikation_1_0.docx

Datum
2015-01-28

Versionsnummer
1.0

Dokumenthistorik

Rev. Nr	Datum	Kommentarer	Ansvarig
0.4	2014-06-19	Text överförd till LA mall och videospecifikation ändrad efter besök hos KB	Leif Pettersson, HiQ
0.5	2014-08-31	Ändringar för leveransformat JPEG och WAVE samt bearbetning av text efter möte 2014-07-02	Leif Pettersson, HiQ
0.6	2014-10-02	Några tillägg efter kommentarer från verksamheten.	Leif Pettersson, HiQ
0.7	2014-10-08	Översyn av kommentarerna från 0.6.	Martin Olsson
0.8	2014-12-22	Bearbetning efter kommentarer från tillsynsenheten.	Martin Olsson
1.0	2015-01-28	Fastställd på förvaltningsledningsmöte.	Martin Olsson

Dokumentnamn
SLL LA Formatspecifikation_1_0.docx

Datum
2015-01-28

Versionsnummer
1.0

1 Inledning

Detta dokument beskriver de filformat som är godkända för leverans av digitala filer till Landstingsarkivet och de förutsättningar som gäller för varje format. Dokumentet försöker ge en så enkel beskrivning som möjligt vad gäller val av format och specifikation för respektive format.

Inga filer kan levereras till Landstingsarkivet separata utan sin kontext. De måste ingå i en leverans med metadata enligt den leveransspecifikation som gäller för leveranstypen. Det finns ett mycket stort antal digitala format för text, bild, ljud och video. Följande format är valda för att de ska möjliggöra att bevara filerna över lång tid.

Vid frågor kontakta Landstingsarkivets förvaltning av e-tjänster för arkivering: [e-arkiv@sll.se](mailto:arkiv@sll.se)

2 Godkända leveransformat

Med godkända leveransformat menas att leveranser till Landstingsarkivet ska, beroende på typ av leveranser, vara i något av de filformat som anges i tabellen nedan. Varje format beskrivs utförligare i kap. 3-7.

Format	Användning och förklaring	Beskrivning	Hänv.
PDF/A	Textbaserade dokument, exempelvis kontorsdokument. Arkivananpassad pdf.	PDF/A-1b (ISO/ICE 19005-1:2005, PDF for Long-term Preservation, Use of PDF 1.4, Level B Conformance), RGB färgkodning. Även PDF/A-1a är tillåtet format för leverans.	3
XML	Textbaserat format. Används enligt schema beskrivet i leveransspecifikation.	XML 1.0 (Extensible Markup Language 1.0, Edition 5, W3C Recommendations utgiven 26 november 2008).	4.1
TEXT	Textbaserad flat datafil	Enligt ISO/IEC 8859-1.	4.2
TIFF	Bildformat, okomprimerad. Används för bildfiler.	TIFF 6.0 UNC RGB (Adobe Systems Inc, TIFF Revision 6.0, 3 juni 1992, Uncompressed, RGB färgkodning).	5.1
JPEG	Bildformat, komprimerad. Om JPEG används ska även TIFF levereras.	JPEG (ISO/ICE 10918-1:1994, Parts 1-6)	5.1.1
BWF	Ljudformat, okomprimerad.	Broadcast WAVE Format, version 1, med LPCM, okomprimerad (EBU Broadcast WAVE Format Technical Specification Tech 3285, version 1, juli 2001). Även Broadcast WAVE Audio File Format 2 är tillåtet format.	6.1
MP3	Ljudformat, komprimerad.	MPEG-1, lager 3 (ISO/IEC 11172-3).	6.1.1

Dokumentnamn

SLL LA Formatspecifikation_1_0.docx

Datum

2015-01-28

Versionsnummer

1.0

Format	Användning och förklaring	Beskrivning	Hänv.
	Om MP3 används ska även Broadcast WAVE levereras.		
MPEG-4	VideofORMAT, används vid konvertering analogt-digitalt eller digitalt-digitalt eller om ursprunglig fil är MPEG-4.	MPEG-4, H.264 (ISO/ICE 14496-14)	7.1
MPEG-1	VideofORMAT, får användas om ursprungliga filen är i detta format.	MPEG-1 (ISO/IEC 11172-2)	7.2
MPEG-2	VideofORMAT, får användas om ursprungliga filen är i detta format.	MPEG-2, H.422(ISO/IEC 13818)	7.3

3 Textbaserade dokument – PDF/A

Format: PDF/A-1b (Portable Document Format/Archive), Level B Conformance. RGB färgkodning. Filändelse ska vara: .pdf

Textbaserade dokument består främst av text men kan även innehålla tabeller, illustrationer och bilder. Det filformat som ska användas för arkivering är PDF/A.

Undantag: Undantaget från den här kategorien är dokument som sedan tidigare har blivit skannade från papper och som har sparats som TIFF-filer.

3.1 PDF/A i korthet

I korthet är en PDF/A-fil en vanlig PDF-fil men den är skapad enligt ett regelverk som begränsar vilka funktioner som får användas. Syftet med formatet är att i så stor utsträckning som möjligt säkerställa att ett dokument som sparats enligt PDF/A-1b ska se likadant ut även efter flera år. De viktigaste kraven är:

- Filen ska vara deklarerad som PDF/A, dvs det ska framgå av filens kod att det är en PDF/A-fil.
- Filen får inte innehålla inbäddade ljud- eller videofiler.
- Filen får inte innehålla JavaScript (skriptspråk, främst använt på klientsidan i webbtillämpningar).
- De fonter (teckensnitt) som används i dokumentet ska vara inbäddade i filen så att filen inte är beroende av extern fontinformation för att återges
- Filen får inte vara krypterad eller lösenordsskyddad.
- Filen får inte hämta data eller information från en extern källa.
- Interaktiva tabeller där utseendet beror på input från extern källa är förbjudna.

För PDF/A-1a tillkommer en rad ytterligare krav. Bland annat:

Dokumentnamn	Datum	Versionsnummer
SLL LA Formatspecifikation_1_0.docx	2015-01-28	1.0

- Dokumentets hierarki ska ingå i filen, till exempel rubrik 1, rubrik 2, etcetera.
- Filen ska vara taggad med bland annat alternativa texter för bilder och illustrationer, alternativa texter för symboler som används i texten som – ”procent” för symbolen ”%”, ”och” för symbolen ”&” samt ”paragraf” för symbolen ”§”.
- Filen ska innehålla Unicode teckenkoder.
- Det eller de språk (exempelvis svenska) som används ska vara angivna i filens kod.

PDF/A-filer kan bestå av konverterade dokument som skapats

- i Microsoft Office-program,
- som e-postmeddelanden,
- när pappershandlingar skannas,
- i andra PDF-format,
- som handlingar som skrivs ut från databaser eller andra verksamhetssystem.

3.2 Konvertera till PDF/A

När en konvertering sker är det viktigt att säkerställa att samtliga delar av ett dokument kommer med i PDF/A-filen, dvs att ingen text går utanför marginalen eller att delar saknas. Det kan vara speciellt utmanande att skriva ut Excelfiler eller filer från andra kalkylprogram. Hela tabellbredden ska få plats på en sida och samtliga flikar av kalkyldokumentet ska vara med i PDF/A-filen om man inte beslutat något annat. Om sidnumrering saknas ska en sådan infogas vid konverteringstillfället.

Vid konvertering av äldre PDF-filer till PDF/A-1 kan funktionalitet försvinna eller förändras. Detta beror på att yngre versioner av PDF-formatet tillåter funktionalitet som inte finns i PDF 1.4, det format som PDF/A-1 bygger på. Bästa sättet att kontrollera detta är att göra en testkonvertering.

Konvertering till PDF/A ska ske så tidigt som möjligt under dokumentets livscykel. Med tiden ökar utmaningen att konvertera eftersom proprietär mjukvara (mjukvara där en ägare satt villkor för användande, spridning och vidareutveckling) och de tillhörande filformaten riskerar att bli obrukbara över tid.

3.3 Transparens

Det finns några utmaningar med att skriva ut eller konvertera till PDF/A som är bra att vara medveten om. PDF/A-formatet tillåter inte transparenta bilder (exempelvis två bilder som går in i varandra och båda syns så att en genomskinlig effekt skapas) vilket kan göra att vissa illustrationer förändras och kan se annorlunda ut jämfört med originalet. Olika diagram med skuggningseffekter, som kan användas i

Dokumentnamn

SLL LA Formatspecifikation_1_0.docx

Datum

2015-01-28

Versionsnummer

1.0

Officedokument, kan vara framställda med den transparenta tekniken och således inte tillåtas av PDF/A-formatet.

3.4 Färgkodning

Vid val av färgkodning ska RGB (Red-Green-Blue) användas. Rekommenderade RGB-format är sRGB (Standard RGB) för vanliga textdokument. Om ett dokument innehåller många fotografier där det är viktigt att bibehålla bra kvalitet på dessa kan Adobe RGB användas.

3.5 Hyperlänkar

Hyperlänkar till externa objekt eller källor är inte förbjudna men har inget garanterat stöd i standarden. Så kallade maskerade länkar, där till exempel ett ord är markerat och hyperlänken/ URL:en anges i bakgrunden, brukar oftast fungera. Det är en rekommendation att en URL anges fullt synlig i texten eller i en not. Då finns alltid möjligheten att manuellt kopiera den och klistra in den i en webbläsare.

3.6 Formulär

Formulär i digitalt format, till exempel Word eller PDF, kan vara mycket svåra att konvertera till PDF/A. Många funktioner som används i formulär hindrar konvertering. Främst är formulär med dynamik, det vill säga att ett val i ett fält styr hur formuläret presenteras och vilken data som kan anges i nästa steg, svåra att konvertera utan speciell programvara. Orsaken är att en grundförutsättning för standarden är att PDF/A-filen ska se ut som grundfilen och med dynamik visas inte hela formuläret.¹

3.7 Tillgänglighetsanpassade PDF

Tillgänglighetsanpassade PDF-dokument enligt standarden för PDF/UA (ISO 14289-1:2012) kan även vara verifierade PDF/A-filer. Det ska i princip även gå att konvertera en PDF/UA-1 fil till PDF/A-1a. Det finns dock vissa funktioner i PDF/UA som inte fungerar i PDF/A-1a. PDF/UA bygger på version PDF 1.7 medan PDF/A-1 bygger på den äldre PDF 1.4. Problem kan uppstå främst om ett dokument innehåller tabeller.

3.8 Skannade handlingar

För skannade handlingar som levereras i PDF/A-format ska följande inställningar göras vid inskanningen. Specifika inställningar för bilder i pdf-dokumentet kan göras.

¹ För att läsa mer om PDF/A och till exempel se möjliga felkoder i Adobe Acrobat, se <http://www.pdfa.org/publication/pdfa-in-a-nutshell-2-0/>

Dokumentnamn	Datum	Versionsnummer
SLL LA Formatspecifikation_1_0.docx	2015-01-28	1.0

Upplösning: 300 PPI (Pixels Per Inch)
Typ av bild: Färg
Komprimering: Utan förlust.
Färg: Rekommendation sRGB (Standard RGB). Se texten ovan.

3.9 Validering

Det är viktigt att validera konverteringar av dokumentfiler till PDF/A-1b eller PDF/A-filer som skapats genom skanning. Enstaka filer kan valideras i till exempel Adobe Acrobat eller 3-Heights™ PDF Validator Online Tool, ett gratisverktyg på webben.² Om större mängder filer ska valideras kan ett serverbaserat verktyg utvärderas.³

4 Textbaserade format

4.1 XML

Format: XML (Extensible Markup Language), 1.0.

Filändelse ska vara: .xml, .dtd eller .xsd

I första hand gäller de leveransspecifikationer som publiceras av Landstingsarkivet för leverans av XML-filer. Andra typer av XML-filer kräver överenskommelse med Landstingsarkivet.

Observera: XML 1.1 är inte tillåten för leverans då det råder osäkerhet om det finns stöd för versionen i alla system. Detta på grund av att nationella specialtecken är tillåtna i tag-texten (etiketterna), det vill säga bokstäverna Å, Ä och Ö.

4.2 TEXT

Format: Enligt ISO/IEC 8859-1. Teckenkodning UTF-8 eller UTF-16 Unicode.

Filändelse ska vara: .txt

Formellt sett är TEXT inget filformat eftersom filen helt saknar formatering. Det rör sig om en flat datafil.

5 Bildformat

Inom den här kategorin handlar det till övervägande del om bilder som har

- fotograferats med digitalkamera
- skannats från fotokopior, diapositiv, negativ eller mikrofilm/mikrofiche.

Om det finns upphovsrättsliga begränsningar är det viktigt att handlingar rörande dessa bifogas eller anges vid leverans.

² Länk: <http://www.pdf-tools.com/pdf/validate-pdf-a-online.aspx> (adress hämtad 2014-08-31)

³ Exempel på verktyg finns redovisade på [pdfa.org:s](http://www.pdfa.org/s) hemsida, <http://www.pdfa.org/2011/08/validating-pdf-a/> (adress hämtad 2014-08-31)

Dokumentnamn
SLL LA Formatspecifikation_1_0.docx

Datum
2015-01-28

Versionsnummer
1.0

5.1 TIFF

Format: TIFF (Tagged Image File Format) version 6.0 uncompressed baseline RGB. Enligt Adobe Systems Incorporated. Utgiven 1992-06-03. Filändelse ska vara: .tif eller .tiff

TIFF är ett rasterbaserat format som har funnits i över 20 år och anses vara mycket stabilt. Vid skanning i nutid av fotografiskt material gäller de specifikationer som anges nedan. För redan befintliga TIFF-filer gäller att dessa ska levereras i okomprimerat format alternativt med CCITT (Comité Consultatif International Téléphonique et Télégraphique) grupp 4-komprimering (förlustfri). Inga andra typer av komprimering är tillåtna utan överenskommelse med Landstingsarkivet.

TIFF-specifikationen tillåter flersidiga TIFF-filer men dessa är svåra att använda. Många bildläsare kan inte se att det finns flera sidor i filen och visar endast den första sidan utan att användaren kan se att det finns flera sidor. Därför ska istället PDF/A användas för flersidiga dokument. Det är emellertid tillåtet att leverera redan befintliga flersidiga TIFF-filer till Landstingsarkivet, skapade innan den här specifikationen blev godkänd

Olika typer av bilder/blad ska skannas med olika upplösning. Upplösningen anges som Pixels Per Inch (PPI). Anledningen till att olika typer av dokument har olika upplösningskrav är att typ av motiv och storlek på originalen spelar in. En A4-sida med svart text på vit bakgrund kan skannas med lägre upplösning än ett 35 mm färgnegativ. Bildfilen av A4-bladet kommer sannolikt att skrivas ut på motsvarande storlek och ha få nyanser. Filen från färgnegativet kommer att förstöras, till exempel till en 10 X 15 cm fotokopia, och ha många nyanser och detaljrikedom.

Nedanstående gäller för:

- dokument med hög kontrast, till exempel textdokument utan rasterade bilder, ritningar, viss grafik, och där eventuell färg inte är värdebärande.
- dokument i färg och viktiga dokument med låg kontrast (till exempel bleknad text, gulnat papper eller med färgbakgrund).

Upplösning	Typ av bild	Bitdjup	Färghantering	Komprimering
Lägst 300 PPI (Pixels Per Inch)	Färg	24-bit	Inbäddad ICC färgprofil	Utan förlust

Nedanstående gäller för papperskopior av svart-vita fotografier och färgfotografier:

Upplösning	Typ av bild	Bitdjup	Färghantering	Komprimering
Lägst 600 PPI (Pixels Per Inch)	Färg	24-bit	Inbäddad ICC färgprofil	Utan förlust

Dokumentnamn

SLL LA Formatspecifikation_1_0.docx

Datum

2015-01-28

Versionsnummer

1.0

Nedanstående gäller för

- transparenta svart-vita fotografier och färgfotografier, till exempel diabilder eller negativ.
- så kallade glasplåtar

Upplösning	Typ av bild	Bitdjup	Färghantering	Komprimering
Lägst 2400 PPI (Pixels Per Inch)	Färg	24-bit	Inbäddad ICC färgprofil	Utan förlust

5.1.1 JPEG-filer

Format: JPEG (Joint Photographic Expert Group) i ISO/ICE 10918 File Information Format (JFIF). Filändelse ska vara: .jpg, .jpeg, .jpe

Bilder i JPEG-format som levereras till Landstingsarkivet ska konverteras till formatet TIFF. Den konverterade TIFF-filen ska ha minst samma upplösning som JPEG-filen. Både den ursprungliga JPEG-filen och TIFF-filen ska levereras till Landstingsarkivet.

JPEG är ett rasterbaserat bildformat som använder en så kallad informationsförstörande algoritm för att komprimera bilder. När det gäller JPEG finns det för många komplicerande faktorer och för stort hårdvaruberoende för att kunna ange exakta formatkrav.

→ Observera! Vid skanning av textdokument är det **inte** tillåtet att använda filformatet JPEG. Komprimeringen gör att texten i dokumenten inte kan teckentolkas och digitaliseras.

→ Observera! Om en bild har blivit sparad i JPEG-format får filen aldrig sparas om på nytt (motsvarande funktionen ”Spara som...”). Om detta sker komprimeras den redan komprimerade bilden en gång till. Det behövs endast några få sådana händelser för att bilden ska få betydligt sämre bildkvalitet. Detta gäller också om en JPEG-bild har behandlats på något sätt, till exempel förstörats, förminskats eller beskurits. I dessa fall ska en kopia av JPEG-filen sparas för att bevara kvalitén i den ursprungliga filen.

→ Om det finns möjlighet att välja upplösning i hårdvaran, t.ex. i digitalkameran eller mobiltelefonen, så ska så stor upplösning som möjligt väljas (t.ex. 4 608 x 3 456 PPI (Pixels Per Inch)). Idealt är om kameran kan hantera RAW-filer och att dessa används för att skapa en TIFF-kopia samt vid behov en separat JPEG-kopia.

6 Ljudformat

Det finns ett stort antal ljudformat i den digitala världen. Följande ljudformat är idag tillåtna för leverans till Landstingsarkivet. Om en ljudinspelning finns i båda angivna formaten ska båda dessa levereras till Landstingsarkivet. Om en analog

LA 2015-0022

Dokumentnamn

SLL LA Formatspecifikation_1_0.docx

Datum

2015-01-28

Versionsnummer

1.0

upptagning konverteras till digitalt format ska varje upptagning konverteras till en Broadcast WAVE Format-fil (BWF) respektive en MP3-fil.

6.1 Broadcast WAVE Format

Format: Broadcast WAVE Format, version 1, med LPCM, okomprimerad, EBU Broadcast WAVE Format Technical Specification Tech 3285, juli 2001. Filändelse ska vara: .wav

BWF-formatet utvecklades av European Broadcasting Union (EBU) för att underlätta utbyte av digitala ljudfiler mellan olika radio- och TV-stationer som hade olika utrustningar. Första versionen av formatet publicerades 1996. Formatet är ett så kallat öppet format. All mjuk- och hårdvara som kan hantera WAVE-filer kan hantera BWF-filer. BWF är en defactostandard inom radio och TV samt börjar väljas som bevarandeformat av allt fler arkivinstitutioner.

BWF-formatet ska användas när konvertering sker från olika typer av upptagningar som ljudband, analoga skivor och andra digitala ljudformat. BWF är ett containerformat, det vill säga en behållare för de faktiska ljudfilerna och andra stödfiler. BWF tillåter endast två ljudfilsformat, MPEG och PCM. Det vanligaste ljudformatet är en version av PCM, LPCM (Linear Puls Code Modulation).

BWF-filen ska vara okomprimerad enligt följande specifikation.

För ljudupptagningar som inte är komplexa, dvs som inte innehåller djupa ljudbilder, till exempel upptagningar med tal:

Samplingsfrekvens: Minst 48 kHz.

Bitdjup: 24 bitar.

För ljudupptagningar med stort djup eller där alla nyanser i upptagningen är viktiga informationsbärare:

Samplingsfrekvens: 96 kHz.

Bitdjup: 24 bitar.

6.1.1 MP3

Format: MPEG-1 (Moving Pictures Experts Group) lager 3. Filändelse ska vara: .mp3

MP3 är främst ett format för tillgängliggörande via Internet. I de fall ljudfiler endast finns i MP3-format ska dessa även konverteras till BWF enligt ovan inför leverans till Landstingsarkivet.

För varje ny BWF-fil, konverterad från annan ljudkälla än MP3-filer, ska en spridningskopia i MP3-format levereras till Landstingsarkivet med nedan angivna specifikationer:

Dokumentnamn	Datum	Versionsnummer
SLL LA Formatspecifikation_1_0.docx	2015-01-28	1.0

Bithastighet, mono: Minst 128 kb/s.
Bithastighet, övriga: Minst 160 kb/s
Samplingshastighet: 48 kHz.
Bitdjup: 16 bitar.

Observera: Om en ljudfil har sparats i MP3-format får filen aldrig sparas om en gång till (motsvarande funktionen ”Spara som...”). Om detta sker komprimeras den redan komprimerade ljudfilen på nytt. Det behövs endast några få sådana händelser för att ljudfilen ska tappa betydligt i ljudkvalitet.

7 Videoformat

Det finns bokstavligen talat hundratals olika filformat för video. Följande tre format är godkända för leverans till Landstingsarkivet. När videoupptagningar konverteras ska MPEG-4 användas i första hand.

7.1 MPEG-4

Format: MPEG-4/H.264 (ISO/IEC 14496-14). Filändelse ska vara: .mp4

Analog video, 4:3 (1,33:1)

Videoström: H.264/AVC, main profile @ L3.1 (ISO/IEC 14496-10), progressiv.

Video bithastighet: 1,5 MB/s (SD - Standard Definition).

CABAC: On

Deinterlace: On

Frame rate: 25 pfs.

Display Aspect Ratio med Storage Aspect Ratio (pixlar):

720 X 576

Ljudström: AAC (ISO/IEC 14496-3).

Ljud bithastighet: 128 kbit/s.

Bitdjup: 16 bitar.

Film och filformat, 16:9 (1,79:1), High Definition

Videoström: H.264/AVC, main profile @ L3.1 (ISO/IEC 14496-10), progressiv.

Video bithastighet: 3 MB/s.

CABAC: On

Frame rate: 25 pfs.

Display Aspect Ratio med Storage Aspect Ratio (pixlar):

16:9 – 1290 x 720. Även 1920 x 1080 är godkänt leveransformat.

Ljudström: AAC (ISO/IEC 14496-3).

Ljud bithastighet: 128 kbit/s.

Bitdjup: 16 bitar.

Om formatförändring sker från 4:3 till 16:9 ska formatet Pillar Box användas med Black Bars. Formatförändring, från 16:9 till 4:3 är inte tillåten.

7.2 MPEG-1

Format: MPEG-1 (ISO/IEC 11172). Filändelse ska vara: .mpg eller .mpeg

LA 2015-0022

Dokumentnamn

SLL LA Formatspecifikation_1_0.docx

Datum

2015-01-28

Versionsnummer

1.0

MPEG-1 är ett godkänt format för leverans av befintliga filer till Landstingsarkivet. Formatet får inte användas för konvertering av videoupptagningar (se avsnittet om MPEG-4 ovan).

7.3 MPEG-2

Format: MPEG-2/H.422 (ISO/IEC 11172). Filändelse ska vara: .mpg eller .mpeg

MPEG-2 är ett godkänt format för leverans av befintliga filer till Landstingsarkivet. Formatet får inte användas för konvertering av videoupptagningar (se avsnittet om MPEG-4 ovan).