

Format för digitalt långtidsbevarande – PDF/A-1

Utvecklingen av den elektroniska förvaltningen ställer nya krav på hantering och bevarande av den information som uppstår i verksamheten. För att möjliggöra ett digitalt långtidsbevarande måste hänsyn tas till bland annat nya tekniska format och lagringsstrukturer. Riksarkivet föreskriver om lämpliga format för framställning och bevarande av allmänna handlingar hos statliga myndigheter.

Allmänna handlingar som är inkomna till eller upprättade hos en myndighet skall bevaras i ursprungligt skick. Detta innebär att de över tiden skall kunna presenteras med samma innehåll och form som de hade när de kom till myndigheten. Alla åtgärder som medför förlust av betydelsebärande data samt möjligheterna att bedöma om en handling är äkta och i ursprungligt skick räknas som gallring. Ett bevarande utan informationsförluster kan möjliggöras om ett lämpligt format används från början. Exempelvis bör konvertering från format som inte lämpar sig för långtidslagring ske då dokumentet färdigställts. De elektroniska handlingarna skall över tiden upprepat kunna presenteras med samma utseende och innehåll som de hade vid tidpunkten för upprättandet. Elektroniska handlingar innehåller metadata och kan också ge information om händelser och beslut under handläggningens gång samt om vad som hänt med handlingen sedan den upprättats. För att denna information skall vara användbar måste handlingarnas autenticitet och integritet säkerställas. Med detta menas att ingen information lagts till, ändrats eller tagits bort.

Elektronisk ärendehantering möjliggör en effektivare förvaltning och en förbättrad insyn i verksamheten för allmänheten. I takt med att allt mer av ärendehantering på myndigheter sker via elektroniska medier, ställs nya krav på hur de handlingar som uppstår ur den elektroniska ärendehantering skall bevaras och vårdas. Problemen och frågetecknen kring digitalt långtidsbevarande är på inget sätt nya, utan har funnits i decennier. Hela tiden tillkommer nya typer av digitala data som skall bevaras, till exempel bilder, video, ljud, webbsidor och kontorsdokument av olika slag. Erfarenheter av att datafiler har blivit oläsbara har gjort att särskilt formatfrågan blivit uppmärksammas. Det finns naturligtvis även andra faktorer som är av betydelse för elektroniska mediers bevarande som exempelvis klimat, databärare och ett säkert förvar. Av lika stor betydelse är tillgången på programvara och hårdvara som kan läsa informationen.

En avgörande faktor för bevarandet av handlingar är att man har nationell och internationell konformitet (likhet och enhetlighet), samt använder öppna, leverantörs- och plattformsoberoende format. Användandet av ett väldefinierat och standardiserat format innebär stora vinster, både ekonomiskt och informationsmässigt, vid eventuell framtida konvertering.

PDF

PDF (Portable Document Format) är ett format som återger dokumentens visuella utseende oavsett om de är ursprungligt upprättade som PDF, konverterade från andra filformat eller digitaliserade från papper eller mikrofilm och oavsett vilken plattform de visas på.

Det finns dock flera orsaker till att PDF inte är det optimala filformatet för långtidslagring. Till exempel finns inga krav på Adobe¹ att publicera specifikationer för kommande versioner av PDF. Vidare kan en PDF-fil ha ett innehåll som inte lämpar sig för långtidslagring, som till exempel kryptering, inbäddade filer och vissa andra typer av applikationer eller ett externt beroende. För att möjliggöra ett bevarande på lång sikt måste det säkerställas att en PDF-fil kan bevaras och presenteras med oförändrat och förutsägbart resultat under lång tid. Detta är det bakomliggande motivet till att arbetet med att ta fram en arkivbeständig PDF-tillämpning påbörjades.

De PDF-specifikationer som finns idag är dock öppna och fria att använda för att tillverka programvaror som stödjer, läser eller producerar PDF. Detta innebär att de specifikationer av formatet som publicerats inte kommer att vara beroende av en specifik leverantör då de kommer att vara tillgängliga också i framtiden. Även olika typer av applikationer som kan användas tillsammans med PDF är fria att producera. Ett exempel är program som gör det möjligt för personer med olika typer av handikapp att ta del av dokumentet.

PDF/A-1

De krav som ställs på att handlingarna skall vara i ursprungligt skick och fullständighet för ett bevarande över tid, måste tas i beaktande vid val av format. PDF/A-1 är utformat för att för vissa typer av dokument uppfylla dessa krav på ett tillfredsställande sätt. Förutom de metadata som finns i dokumentet finns även möjligheter att förse dokumentet med ytterligare metadata. Detta gör dokumentet spårbart och visar dess historik. Man kan finna information om vem som upprättat dokumentet och när detta skett samt om någon ändring skett i dokumentet eller i metadata.

PDF/A-1 är systemoberoende vilket innebär att dokumentet presenteras med det innehåll och med den layout det hade vid upprättandet, oberoende av den applikation dokumentet framställdes i. Vidare säkerställer PDF/A-1-formatet att filen innehåller all den information som krävs för en säker och förutsägbar tolkning av data över tiden. En annan viktig egenskap hos formatet är att filen är självdokumenterande, det vill säga att den innehåller metadata om innehåll, struktur, tekniska egenskaper och administrativ historik.

¹ Adobe är det mjukvaruföretag som ursprungligen ligger bakom PDF-formatet och som valt att göra specifikationen fritt tillgänglig för andra företag att använda.

Standarden

PDF/A-1 definierar ett filformat baserat på PDF som möjliggör representation av elektroniska dokument så att också layouten bevaras över tiden, oberoende av de program och verktyg som användes för att framställa det ursprungliga dokumentet. Utvecklingen av en PDF/A-1 initierades av det Administrativa kansliet för *U. S. Courts*². 2002 bildades en ISO-kommitté bestående av arkivinstitutioner och bibliotek som NARA (National Archives and Records Administration), Harvard University, olika intresseorganisationer, privata intressenter, Adobe och andra mjukvaruleverantörer. Kommitténs arbete bedrevs under två amerikanska standardiseringsorganisationer, NPES (Association for Suppliers of Printing, Publishing and Converting Technologies)³ och AIIM (Enterprise Content Management Association)⁴. Syftet med arbetet var att definiera ett format för långtidslagring som bevarade dokumentets statiska utseende, innehöll ett ramverk för att lagra metadata samt ett ramverk för att definiera den logiska strukturen.

PDF/A-1 blev en ISO-standard⁵ den 13 september 2005 och svensk standard i oktober 2005. Standarden beskriver hur PDF-specifikationen skall tillämpas för att arkivkraven skall kunna tillgodoses. Godkännandet av PDF/A-1 som en ISO-standard och format för långtidslagring är ett steg på vägen i arbetet att uppnå konformitet gällande tillvägagångssätt och format för bevarande av elektroniska dokument. Standarden utgör ett ramverk för att också kunna bevara de metadata som ger en beskrivning av dokumentets historia och sammanhang.

Ett bevarande över tid säkerställs inte enbart av valet av lämpliga format. PDF/A-1 ensam säkerställer alltså inte att det ursprungliga utseendet överensstämmer med originaldokumentet eller arbetsdokumentet, utan utgör bara en del av den bevarandestrategi inom en organisation som är nödvändig för ett långtidsbevarande av elektroniska dokument. Andra faktorer som kan påverka säkerställandet är exempelvis informationssäkerhet, klimat, databärare, hantering och migrering.

Standardens disposition

PDF/A-standardens beskriver en delmängd av PDF-formatet med endast vissa utvalda objekt eller funktioner som baseras på specifikationen för PDF 1.4⁶. Standarden specificerar hur PDF 1.4-specifikationen får användas för att uppfylla kraven för PDF/A-1. Standarden delar in funktionerna i specifikationen

² U.S. Courts – De amerikanska federala domstolarna: Högsta domstolen, de appellationsdomstolarna, de regionala federala domstolarna samt konkursdomstolarna.

³ NPES – <http://www.npes.org>, <http://www.npes.org/standards/toolspdfa.html>

⁴ AIIM – <http://www.aiim.org>

⁵ SS-ISO 19005-1 Dokumenthantering – Filformat för elektroniska dokument för långtidsbevarande – Del 1: användning av PDF 1.4. Standarden kan köpas från sis – www.sis.se

⁶ PDF Reference version 1.4 – http://partners.adobe.com/public/developer/pdf/index_reference.html

i tvingande, frivilliga eller förbjudna samt specificerar hur dessa får *användas* i mjukvara för att framställa pdf-filen.

Standarden listar normativa referenser, till exempel *ExtensibleMarkupLanguage (XML) 1.0 (Second Edition)*, *W3C Recommendation*, *PDF Reference: Adobe Portable Document Format, Version 1.4 3rd ed*, vilka genom detta alltså utgör förutsättningar för PDF/A-1.

Paragraferna i kapitlet om de tekniska kraven beskriver tvingande, frivilliga eller förbjudna egenskaper hos filen. Kapitlet inleds med en paragraf om filformat som specificerar utmärkande egenskaper hos pdf-filen, nödvändiga för att producera enhetliga PDF/A-filer. Därefter följer paragrafer som rör grafik, teckensnitt, transparens, kommentarer, händelser, metadata, logisk struktur och formulär i dokumentet.

Kortfattat gäller att rendering⁷ av grafik skall vara oberoende av applikationer utanför filen, att teckensnitten är inbäddade och att metadata finns i XMP (eXtensible Metadata Platform)⁸. Transparens⁹ i dokumentet är inte tillåtet, inte heller externa länkar, inbäddade exekverbara filer, script, externa beroenden av något slag eller formulärfält innehållande funktioner.

XMP – Metadata

PDF-dokument innehåller metadata som rör dokumentegenskaperna, men begränsas inte till dessa. Adobe har utvecklat XMP som är en del av PDF/A-1-standard och som utgör ett sätt att organisera och beskriva metadata. Genom att förse filerna med metadata kan de organiseras och användas på ett effektivare sätt, exempelvis genom förbättrad sökning både av dokumentets egenskaper och av dess innehåll. Den information som utgör metadata kan inhämtas och integreras i dokumentet då det skapas, och samtidigt föras in i organisationens dokumenthanteringssystem.

Specifikationen av XMP är liksom för PDF öppen, baserad på W3C-rekommendationer¹⁰, uttryckt i XML och följer även *Dublin Core Initiative*¹¹, en internationell metadatastandard. Detta gör att utvecklare har möjlighet att implementera eller anpassa XMP i olika produkter som används för att skapa lösningar för dokumenthantering anpassade till en specifik verksamhet.

⁷ Rendering: *Återge eller rita upp*. När en dator "renderar" skapar den en bild som kan visas på bildskärmen. Ordet används om krävande bilder där datorn måste räkna ut skuggor, reflexer och perspektiv.

(<http://computersweden.idg.se>)

⁸ XMP – <http://www.adobe.com/products/xmp/>

⁹ Transparens – genomskinlighet. Innebär att olika lager kan skina igenom varandra, används exempelvis vid dimeffekt m.m.

¹⁰ W3C – World Wide Web Consortium, ett internationellt konsortium som utarbetar rekommendationer och vägledningar för Internet med målet att uppnå interoperabilitet mellan olika hårdvaror och programvaror på webben. <http://www.w3.org>

¹¹ Dublin Core initiative – <http://www.dublincore.org>, ISO-standard heter "ISO 15836:2003 Information and documentation – The Dublin Core metadata element set"

PDF/A-2

En ny del i standarden utarbetas för närvarande. Denna bygger på pdf-specifikationen 1.6 och kommer således att innehålla en del av de funktioner som tillkommit i denna, exempelvis digitala signaturer. PDF/A-2 kommer inte att påverka filer som uppfyller kraven för PDF/A-1, däremot kommer inte PDF/A-2-filer nödvändigtvis uppfylla kraven för PDF/A-1.

Varför PDF/A?

Riksarkivet föreskriver inte om användning av en specifik produkt eftersom ett bevarande kräver ett oberoende av leverantörsspecifika produkter. I första hand hänvisas till internationell standard, därefter övriga standarder och i sista hand de tekniska krav som Riksarkivet utarbetat. PDF/A-1 är ett format som möjliggör ett bevarande i ursprungligt skick för vissa typer av dokument. Formatet är inte avsett att användas för att framställa och redigera dokument utan för att upprättandet av den slutliga handlingen.

Formatet lämpar sig för upprättande av kontorsdokument efter det att det färdigställts i exempelvis ordbehandlingsprogram. Även e-post som skall bevaras kan sparas som PDF/A.

Att rent praktiskt framställa PDF/A-1 behöver inte bli besvärligare än att framställa en PDF. Idag finns det programvara som stödjer standarden och skapar dokument som följer standarden för PDF/A-1 och som också kan kontrollera om färdiga PDF:er uppfyller kraven för PDF/A-1. Inom en framtid kan flera nya programvaror med funktioner för PDF/A-1 väntas på marknaden som en effekt av godkännandet av standarden och det faktum att koden är fri att använda för alla mjukvaruleverantörer.

För ytterligare information, kontakta Torbjörn Hörnfeldt på fornamn.efternamn@riksarkivet.ra.se