

PDF: Fleksibelt, fritt, tilgjengelig

Om noe skulle fortjene betegnelsen 'en gammel kjenning' i dagens IT-hverdag, må det være PDF. Den synes å ha vært der så lenge vi kan huske, og har reddet oss så mang en gang når andre formater har skapt kompatibilitetsproblemer. Og den har vært en selvfølge for materiale som skal overføres til trykkeri eller grafiske eksperter i forbindelse med alle slags trykksaker.

Vårt digitale papir

Sammen med Internett-eksplosjonen har Adobes PDF styrket sin rolle som universalformat for formidling og presentasjon av all slags informasjon som tidligere ble trykket, men som i dag blir distribuert digitalt. Dette må kunne kalles en suksess, men hva er årsaken til suksessen og hva er det PDF har som har gjort formatet til slik en selvfølge? Og hvorfor påkaller PDF vår oppmerksomhet utover det faktum at den er en allesteds nærværende hjelper i hverdagen? Dessuten – i en tid med voksende fokus på åpenhet og tilgjengelighet, hvor korrekt er overskriften ovenfor? Hvor fritt og tilgjengelig er PDF, og ikke minst hvor pålitelig er formatet som standard? Kan vi om 20 år gå tilbake til dagens dokumenter og fortsatt lese dem? Armert med svært så lite oppløftende erfaringer fra Microsofts Word-format blir ikke minst det siste et viktig spørsmål.

Bakgrunn

Vi skal ta for oss disse og flere spørsmål nedenfor, og begynner med årsaken til suksessen. Historien går helt tilbake til Postscripts fødsel, som var tett knyttet til introduksjonen av Apples Macintosh i 1984, og forårsaket en revolusjon for *desktop publishing*. Fonter – skrifttyper – er en essensiell komponent i slike systemer, og må finnes både på systemene som produserer dokumenter og på enheten som skal skrive dem ut. Alternativt kan fontene inkluderes – legges ved – når en Postscript-fil genereres for utskrift.

Her starter imidlertid komplikasjonene. For det første er skrifttyper lisensiert til brukeren, som normalt har beskjedne rettigheter til å flytte dem utenfor eget system. Dessuten – etterhvert som Postscripts popularitet spredte seg til andre plattformer, ble nye fontformater tatt i bruk – med påfølgende kompatibilitetsproblemer. Dokumentutveksling ble vanskelig fordi PCer og Macintoshen ikke hadde de samme fontene, og overføring av fonter mellom systemene lot seg ikke gjøre verken av lisensmessige eller tekniske årsaker.

Som om ikke dette var nok, oppdaget Microsoft allerede før 1990 at Postscript var i ferd med å innta en rolle selskapet helst ville kontrollere selv. Dermed ble det i Windows lagt kjepper i hjulene for Postscript, hvilket kompliserte både font-problematikken og andre forhold.¹

Postscript hadde imidlertid allerede satt seg fast i trykkeribransjen, med printere, fotosettere og annet utstyr som hadde byttet ut utallige proprietære dataformater med Postscript. Toget var gått, men noe måtte gjøres for å forenkle fontproblematikken og dessuten få kontroll med de enorme filene som ble generert. Selv uten bidraget fra Microsoft, var filstørrelse en betydelig utfordring fordi utskriftsfilene inneholdt bilder og grafikk av høy oppløsning.

Til hjelp i font-kaos

PDF ble utviklet av Adobe tidlig på 90-tallet for å løse disse problemene. Font-problematikken ble ordnet på den måten at fontene inkluderes i dokumentet (PDF-filen), men ikke i sin helhet, og kodet på en måte som gjør det umulig å separere dem. Dermed var lisensproblematikken under kontroll. Videre ble selve dokumentkoden langt mer effektiv enn Postscript, takket være et mer kompakt kodingsformat og ikke minst via intelligens i konverteringen fra Postscript til PDF. Acrobat Distiller, som verktøyet heter, sørger – i tillegg til å komprimere objektene mest mulig – for å fjerne all usynlig informasjon. Mengden som dermed blir borte, kan anta anselige proporsjoner i dokumenter med mye grafikk. Et stort og detaljert bilde – hvorav kun enkelte deler er synlige, mens det meste havner bak andre objekter, er et nærliggende eksempel. Størrelsen på en PDF-fil er derfor gjerne på mellom 2 og 10% av den opprinnelige Postscript-filen. PDF er mer kompakt enn de fleste andre dokumentlagringsformater som er i bruk i dag.

En Acrobat inntar scenen

Sammen med det nye formatet kom PDF-leseren Acrobat Reader, som den første tiden kostet USD 50 og var tilgjengelig kun for Mac og siden Windows. Denne prisstrategien fungerte åpenbart ikke, og leseren – som i dag heter Adobe Reader – ble gratis fra høsten 1994.

Populariteten ble imidlertid fortsatt hemmet av stiv pris og trange lisensbetingelser for konverteringsprogrammet (*Distiller*). Det er i ettertid lett å se at Adobe både underestimerte PDFs potensiale, og brukte for lang tid til å bringe konverteringskostnadene ned til et akseptabelt nivå for både store og små organisasjoner.² Hadde selskapet spilt sine kort litt annerledes i denne perioden (1995-2000), ville PDF ha hatt en vesentlig større rolle – blant annet som arkivformat for dokumenter. I stedet spilte

Historiske pussigheter

PDF ble første gang nevnt offentlig på en *publishing*-konferanse i San Jose i 1991, den gang med betegnelsen IPS, *Interchange PostScript*.

Høsten etter ble teknologien i sin helhet presentert på Comdex-konferansen, og høstet mange lovord – og premiering som beste teknologi. Sommeren 1993 kom produktene på markedet. Mens anvendeligheten allerede da var god for utveksling og utskrift av dokumenter, var første generasjon uinteressant for grafisk bransje. Av ukjente og uforståelige årsaker hadde Adobe utelatt både CMYK-fargemodellen og andre elementer som bransjen var helt avhengig av.

Høsten 1994 reiste Adobe kjerringa – med Acrobat 2, PDF 1.1, støtte for flere fargemodeller, og en lang rekke andre forbedringer som spesielt grafisk bransje var avhengig av. Samtidig ble Acrobat Reader gratis – og sluppet for flere plattformer.

- 1 Vi sliter fortsatt med konsekvensene av disse 'kjeppene' fra Microsoft når Windows-programmer skal generere utskrifter til Postscript-enheter. Utskriftsfilene som produseres, blir gjerne bortimot 10 ganger større enn nødvendig, fordi de må gjennom Windows' utskriftssystem, som – overlatt eller ikke – er ekstremt ineffektivt.
- 2 Vi foretok en analyse av disse lisens-kostnadene for en klient med over 10.000 brukere i 1998/99, og kunne – selv etter forhandlinger med Adobe – konstatere at utgiften per bruker var en størrelsesorden utenfor hva som kunne forsvares.

Adobe denne ballen rett i hendene på Microsoft, og la grunnlaget for den sterke posisjon Word-formatet (.doc) har i dag.

Internett som katalysator

Først da prisen på PDF-leseren ble redusert til null og den ble gjort tilgjengelig på flere plattformer, ble det fart i sakene. PDF 1.2, som kom til verden i 1996, løste akutte problemer for et sultent marked, i og utenfor grafisk bransje. Tilgjengelighet, plattformuavhengighet, samtidighet med Internett-ekspløsjonen og ikke minst en 'plug-in'-utgave av Acrobat Reader for Netscapes nettleser, la grunnlag for en spredning som virkelig monnet.

Det er på mange måter ironisk at Internettet og Web-teknologien ble katalysatorer for en teknologi med en så bunnsolid forankring i en tradisjonell, papirbasert verden som PDF. Vanskelighetene med å få web-sider til å gi en forutsigbar og tiltalende presentasjon av innhold, som karakteriserte WWW inntil CSS-teknologien kom på banen, har en god del av æren for PDFs gjennombrudd.

CSS – Cascading Style Sheets, se www.w3c.org for detaljer.

Fra presentasjon til intelligens

Allerede med versjon 1.2 var PDF blitt langt mer enn en avansert form for Postscript. Definisjoner for metadata som innholdsfortegnelser, eksterne referanser, notater og sikkerhetsmekanismer var på plass, og utvidelsene fortsetter. Adobe har åpenbart ambisjoner for PDF og tilhørende verktøy som strekker seg langt utenfor teknologiens opprinnelige nedslagsfelt – på godt og vondt.

Dermed nærmer vi oss årsaken til at PDF i dag er interessant i langt større sammenhenger enn opprinnelig tiltenkt. For det første har PDF allerede funnet veien til en rekke anvendelser der fontkompatibilitet er en marginal ingrediens. Dernest er behovet for mer avanserte mekanismer i forbindelse med digital distribusjon av dokumenter mer påtrengende enn noen sinne – ikke minst i lys av sterkere fokusering på opphavsrettigheter. Og for det tredje har PDF – eller rettere sagt Adobe – gang på gang demonstrert formatets fleksibilitet med hensyn til utvidelser. For eksempel vakte det ikke så rent lite oppsikt da illus-

trasjons/tegne-programmet Illustrator i versjon 7 kunne lagre dokumentene i PDF-format. At verktøyet kunne generere PDF-filer var en selvfølge, men å bruke PDF som lagringsformat for hele dokumentene, med metadata, strukturinformasjon og andre data som hører hjemme i kompliserte, redigerbare dokumenter, var noe nytt.

Adobe beviste på den måten at PDF ikke bare var et format for presentasjon og distribusjon av ferdige dokumenter, men også kunne brukes som generisk lagringsformat. Konseptet er egentlig svært enkelt, og er det samme som

Hva er PDF/X?

Selv etter at PDF 1.2 kom på banen, forble formatet et smertens barn for grafisk industri. Spesifikasjonen var så åpen – løs i kantene – at helt legitime PDF-filer fra et voksende antall verktøy var umulige å håndtere. Samtidig innså bransjen at PDF hadde det beste potensiale til å løse de voksende utfordringene med formater, fonter og enorme datamengder.

Derfor tok bransjen skjeen i egen hånd, og startet i 1997 arbeid med en egen standard basert på PDF 1.2. Standarden ble ferdig i 1999 og registrert som offisiell amerikansk standard. Året etter ble PDF/X-1 tatt i praktisk bruk, og er siden oppdatert flere ganger – PDF/X-1a, PDF/X-2 og PDF/X-3, som er under arbeid.

Det er først og fremst i forbindelse med annonser disse standardene har fått stor praktisk betydning, for aviser, annonsebyråer og reklamebyråer. For det vi kan kalle alminnelig bruk, med langt enklere krav på alle nivåer, er Adobes PDF – som nå befinner seg på versjon 1.6, mer enn rikelig.

benyttes i web- (HTML-) sider: Kodene som ikke har med selve visningen å gjøre, kapsles inn slik at de blir ignorert av verktøy som ikke forstår noe av dem. Dermed kan en PDF-leser (som Adobe Reader) fortsatt lese og vise en Illustrator-fil som er full av det vi kan kalle 'annen informasjon'.

Fleksibilitet og utvidelser

Denne fleksibiliteten er årsaken til at PDF i dag står på agendaen i nye sammenhenger. Publisering og utveksling av dokumenter foregår med større hyppighet og i større volumer en noen sinne, mens papirets deltagelse i ligningen er fallende. Digital distribusjon er for mange typer informasjon blitt regelen i stedet for unntaket, hvilket forårsaker nye utfordringer – for eksempel i forbindelse med opphavsrettigheter.

Dagens PDF-format – kombinert med verktøy fra Adobe eller andre leverandører som har spesialisert seg på nettopp PDF – tilbyr mekanismer for blant annet:

- ✓ Sikring av dokumenter: Innholdet kan krypteres og utrustes med passord. Det kan også sperres for utskrift, lokal lagring og annen lokal manipulasjon. Også mulighet for klipp og lim fra PDF-dokumenter kan blokkeres.
- ✓ En PDF-fil kan inneholde programmer – i JavaScript, hvilket på den ene siden gir enorm fleksibilitet, mens det på den andre siden representerer en betydelig sikkerhetsrisiko. PDF-filer kan inneholde virus, og slike virus har forekommet. Som tilfellet er med andre formater med slike egenskaper, er det imidlertid leseverktøyet og i hvilken grad de interne programmene kan aksessere objekter utenfor dokumentet, som avgjør graden av risiko. Likevel, sikkerhetsrisikoen har sørget for at denne egenskapen er sjelden brukt i PDF.
- ✓ Referanser: Dokumentene kan inneholde innkapslet multimedia-innhold, eller eksterne linker til slikt innhold.
- ✓ Arbeidsflyt: Adobe posisjonerer PDF og Acrobat-produktene som verktøy i forbindelse med arbeidsflyt-løsninger. Utgangspunktet er bra, men det er ikke like innlysende hvordan dette passer inn i en utvikling mot Web-tjenester.

Et annet poeng – som egentlig drar i en helt annen retning, er Apples satsing på PDF som innebygget element i brukergrensesnittet i Mac OS X. Med PDF-støtte på dette nivå oppnår Apple å gjøre visning av de fleste tenkelige PDF-dokumenttyper til en triviell affære, og skaper et miljø som blir spesielt PDF-vennlig. Alle verktøy kan implisitt generere PDF-filer, utskrifts-systemet blir homogent og agnostisk med hensyn til formater (PDF kan skrives ut til en hvilken som helst printer) og så videre.

For Apple og Mac-brukere fremstår denne koblingen som effektiv, enkel og elegant. Adobe, som har lisensiert bibliotekene til Apple, får på sin side en praktisk demonstrasjon på anvendelsesbredden til PDF.

Oppsummering

Den kraftige ekspansjonen i anvendelsesområder og funksjonalitet for PDF, kan i første omgang virke forvirrende: Hva er egentlig PDF i ferd med å bli? Og kan ikke alle disse utvidelsene ødelegge den pålitelighet og effektivitet som har gitt PDF den posisjon teknologien har i dag? Risikoen kan ikke avvises. Om vi har fått aldri så mye mer og rimeligere ressurser – lagring, prosessor, båndbredde og så videre – de siste årene, er plass- og kommunikasjonsmessig effektivitet fortsatt viktige attributter.

Uansett er det åpenbart at Adobe har voksende ambisjoner for denne gullfuglen, som uten tvil er det viktigste som har kommet fra selskapet siden Postscript – både teknologisk og forretningsmessig. Mens disse ambisjonene kan få konsekvenser for vår oppfatning av teknologien i fremtiden, forblir PDF for de fleste av oss et lagringsformat og en mekanisme for distribusjon og presentasjon av informasjon. Som sådan har PDF ingen konkurranse – verken fra HTML eller XML. Sistnevnte har riktignok lenge vært ansett som en trussel for PDF på sikt, hvilket ikke kan utelukkes. Denne 'sikten' er imidlertid for lang til å være relevant for praktiske evalueringer i dag.

De funksjonelle utvidelsene, spesielt de som er knyttet til sikring og håndheving av opphavsrettigheter, øker PDFs attraktivitet både i eksisterende segmenter og i nye sammenhenger. Formatets åpenhet fjerner risikoen for å bli låst inn av én spesiell leverandør og bidrar til konkurranse og kreativitet. Og sist men ikke minst har PDF gjennom 10 år demonstrert pålitelighet på tvers av generasjoner. Om våre dokumenter er 5, 6 eller 10 år gamle, kan vi fortsatt lese dem med dagens verktøy. Denne

egenskapen gjør PDF til en velkvalifisert kandidat for langtidsarkivering av data. Muligheten for å sperre innholdet mot endring gjør det samme.

Sammen vil disse egenskapene sørge for at PDF blir et enda viktigere element i vår IT-fremtid enn i dag. ■

Åpenhet, tilgjengelighet og standardisering

Det hele virker unektelig noe kaotisk. PDF er åpent, men eies og videreutvikles av Adobe. Kompatibilitet og plattformuavhengighet er blant formatets viktigste fortrinn, men det finnes mer enn en håndfull versjoner – innenfor og utenfor Adobes kontroll. Og apropos tilgjengelighet – hvordan er det med lisensiering?

For å begynne med det siste, trengs det ingen lisensiering og følgelig påløper det ingen avgift for å bruke PDF-formatet. Spesifikasjonene er åpent tilgjengelige, og hvem som helst kan lage programmer som genererer eller leser formatet. Lisenser kommer først inn dersom programvare fra Adobe benyttes. Selskapet tilbyr en lang rekke tjenester knyttet til PDF, inklusive biblioteker for lesing og skriving av formatet. Dette er ingen triviell oppgave, og svært mange, trolig de fleste programvareleverandører som behandler PDF på en eller annen måte, lisensierer PDF-bibliotekene fra Adobe. Det finnes også Open Source-alternativer, med Ghostscript som det primære eksemplet (www.ghostscript.com).

Med hensyn til kompatibilitet er bildet nyansert. Hvorvidt 'gamle' PDF-dokumenter er lesbare kommer utelukkende an på verktøyet som benyttes. Adobes Reader (gratis) kan lese PDF-dokumenter uavhengig av alder, og de fleste leseverktøy vi har testet kan gjøre det samme. Samtidig er det et faktum at stabiliteten hos verktøyene er variabel. Og sist, men ikke minst er i alle fall Adobe Reader fleksibel med lesing av versjoner som er nyere enn leseren. Den sier ganske enkelt fra om at dokumentet kan inneholde funksjonalitet eller innhold som denne leseren ikke har forutsetninger for å forstå.

Når det gjelder plattformuavhengighet er situasjonen heller ikke 100% entydig. De fleste PDF-dokumenter som ikke inneholder spesielle funksjoner eller restriksjoner, kan leses av Adobe Reader eller andre lesere på de fleste tenkelige plattformer – fra PalmOS, PocketPC og Symbian (mobiltelefoner, PDAer) via Windows, Mac og Linux til samtlige Unix-varianter av markedsmessig betydning.