



# Om at konvertere PDF

- den gode, den dårlige og den forfærdelige metode

<b>Forfatter</b>	<b>Leonard Rosenthal</b> PDF Standards Architect, Adobe Inc.
<b>Oversættelse</b>	<b>Søren Frederiksen / Søren Winsløw</b> DDPFF – <a href="http://www.ddpff.dk">www.ddpff.dk</a> – <a href="http://www.kankanikke.dk">www.kankanikke.dk</a>
<b>Dato</b>	25. oktober 2008
<b>Dato</b>	Oversættelse: august 2009
<b>Status</b>	Endelig



# Indhold

Indhold .....	2
Baggrund .....	3
Terminologi .....	3
Hvordan dobbelt-konvertering kan gå galt .....	4
• Skifter og tekst.....	4
• Gen-komprimering .....	4
• Bløde skygger .....	5
• Farvestyring .....	5
Hvad skal man gøre? .....	5
Alternativer til dobbelt-konvertering .....	6
Konklusion .....	6



## Baggrund

Da Adobe skabte PostScript, var det designet som et printer-specifikt sprog til beskrivelse af indhold af hver side, der skal fortolkes i en printer/rip. I dag kalder vi sådanne sprog PDL (Page Definition Languages eller sidebeskrivessprog), og de bruges fortsat til at styre printere, ripper og lignende. PostScript, der ikke er blevet opdateret siden Level 3 i 1998, definerer en fælles kerne af sprog-egenskaber. Imidlertid bruger hver printer forskellige ekstra udvidelser, der understøtter deres specielle behov – og de er ikke nødvendigvis kompatible med andre printerenheder.

Som tiden gik, og eftersom PostScript blev en standardmetode til at producere og udveksle filer i prepress og trykindustrien, blev det nødvendigt at definere en enheds-uafhængig fælles 'kerne' af PostScript, der kunne bruges til at beskrive en enkelt sides (eller del af en sides) indhold med henblik specifikt på udveksling af grafik, annoncer og lignende. Det er/var, hvad vi kalder EPS (indkapslet (encapsulated) PostScript). Og mange bruger fortsat EPS som en del af deres workflow.

Eftersom PDF originalt delte en fælles fortolkningsmodel med PostScript, var det fornuftigt af Adobe at skabe et værktøj, der konverterer PostScript (PS) og EPS til PDF – Acrobat Distiller, som har været til rådighed siden Acrobat 1.0. Imidlertid har PDFs sidebeskrivelsesmodel udviklet sig væk fra PS-sidebeskrivelsesmodellen – først med introduktionen af transparens fra og med PDF 1.4 og sidenhen med ekstra muligheder som fx JPEG2000 komprimering og lag (Optional content), som bliver brugt mere og mere i trykindustrien.

Dette betyder, at mens produktion af PDF på basis af PostScript fortsat er en acceptabel metode – så er konvertering af en PDF tilbage til PostScript eller EPS ikke nødvendigvis acceptabelt – det kan være en metode med risiko for store tab, og kan resultere i mange problemer senere i workflowet.

Dette dokument har til formål at beskrive nogle af de forskellige problemer, der kan opstå, og beskrive, hvad du kan gøre for at forsøge at undgå problemer, såfremt du er nødt til at konvertere fra PDF til PS / EPS.

## Terminologi

Mens det er uklart, hvorfor det nu hedder, som det hedder, så er det blevet sådan, at den engelske term refrying (på dansk 'at dobbelt-konvertere') har udviklet sig til at betyde processen at konvertere en pdf-fil til PostScript (eller EPS) – og derefter tilbage til PDF.

Man kan foretage denne handling eksplicit, som fx ved at bruge Adobe Acrobat til først at "gemme som PostScript" og derefter at køre denne ps-fil igennem Distiller, eller mere almindeligt, printe til 'Adobe PDF printer'. Denne praksis er startet år tilbage som en metode til at konvertere 'problematisk pdf-filer' til pdf-filer, der var i stand til at printe korrekt. Mens der i dag findes moderne værktøjer til at håndtere sådanne problemer direkte i pdf-filen – er der mange brugere, der fortsat insisterer på at bruge denne metode, fordi "det virkede engang, så det er sikkert også en god metode i dag".

Derudover arbejder man muligvis i applikationer som fx Quark, Adobe PageMaker eller Adobe FrameMaker, der altid vil konvertere en placeret PDF til EPS. Når det færdige dokument derefter printes til 'Adobe PDF printer' genereres der en PostScript-fil, som



derefter sendes til Distiller, og som en konsekvens heraf er en refrying / dobbelt-konvertering faktisk blevet gennemført.

Der er også en anden proces, der nogle gange også kaldes refrying / dobbelt-konvertering, der imidlertid mere præcist bør kaldes 'gen-placering'. At genplacere en PDF er at bruge den på en måde, der ikke (nødvendigvis) var den originale intention. Placering af en PDF af en annonce i en større publikation er et eksempel på en gen-placering, et andet eksempel er udskydning af flere dokumenter i et nyt dokument-layout.

I nærværende dokument vil vi imidlertid udelukkende fokusere på refrying / dobbelt-konvertering og efterlade diskussionen om gen-placering til et andet dokument, der alene dækker dette emne.

## Hvordan dobbelt-konvertering kan gå galt

Som tidligere nævnt involverer refrying / dobbelt-konvertering en dobbelt konvertering – først fra PDF til PostScript og derefter tilbage til PDF. Når man arbejder i et PostScript/EPS baseret workflow, som fx Quark, kan man endda være udsat for op til 4-8 konverteringer af data. Ved enhver transformering af data er der naturligvis risiko for fejl og mistede data.

- **Skifter og tekst**

Skrifterne i Postscript og PDF er omtrent identiske, i hvert fald set ud fra et print synspunkt (og det var jo det Postscript var designet til). Men PDF rummer en række andre features, der fx giver en mere sikker fortolkning af glyffer og giver mulighed for at trække teksten tilbage, ud af PDFen i en klar og stringent form, typisk som Unicode. En typisk konsekvens af dobbelt-konvertering er derfor, at man ikke længere kan trække teksten ud af dokumentet, selv ikke ved en simpel copy-and-paste. Dette har også indflydelse i forhold til søgemaskiner og disses evne til at fritekst-søge i PDF, inkluderet dem bygget ind i operativsystemet, der således ikke mere kan fortage deres job, at indexere indholdet.

I trykverdenen er det ganske almindeligt for et trykkeri at have, hvad kaldes 'sidste øjeblikks rettelse' – hvor en slåfejl lokaliseres og skal redigeres direkte i PDFen – hvilket ses som nemmere end at gå tilbage til det oprindelige dokument. Når skrift-information gennemgår refrying / dobbelt-konvertering, er der en risiko for, at skifterne bliver konverteret i en myriade af metoder, som ganske simpelt kan have indflydelse på muligheden for at redigere pdf-filen på denne måde.

- **Gen-komprimering**

Eftersom PDF supporterer billedkomprimeringsteknologier som fx JPEG2000 og JBIG2, som PostScript ikke gør, så kræver en konvertingsproces, at billeddata bliver de-komprimeret og (muligvis) re-komprimeres med en anden algoritme, som fx JPEG eller CCITTG4. PostScript algoritmerne er ikke så effektive som PDFs algoritmer, så det er forventeligt, at datamængden forøges betragteligt i denne proces. Selv hvis pdf-filen blot



bruger JPEG, så er det sandsynligt, at ukorrekte indstillinger såvel i eksporten til PostScript (EPS) som konverteringen via Distiller til PDF kan resultere i databas.

Ud over nyere komprimeringsalgoritmer, så supporterer PDF 16-bits-billeder, mens PostScript blot supporterer 8-bit. Her er et andet emne, der resulterer i billed-databas, eftersom halvdelen af hver pixel i billedet må fjernes.

- **Bløde skygger**

PostScript Level 3 introducerede et indbygget sæt af operatører for fortolkning af bløde skygger (også kendt som forløb) – og disse blev også introduceret i PDF 1.3. Konsekvensen af dette er, at en hvilken som helst konvertering fra PDF til PostScript/EPS med brug af bløde skygger skal foretages med brug af Level 3 – og ikke den ældre Level 2. Hvis konverteringen foretages med brug af Level 2, så vil den bløde skygge blive konverteret til en serie af linier eller stier (paths), som vil resultere i en enorm forøgelse af filstørrelsen og reducere kvaliteten af fortolkningen, i kraft af striber (banding) og ikke lineære farver. Derudover, eftersom objektet ikke vil være ét objekt, men flere, så er evnen til at lave 'sidste øjeblik-ændringer' også her elimineret.

- **Farvestyring**

Eftersom PostScript var designet med henblik på fremstilling på en præ-defineret enhed (printer) (enhedsafhængigt workflow) – så er al farve også specificeret i denne sammenhæng – og tilbyder dermed ikke de rige muligheder for enhedsafhængige farverum, som er til stede i PDF – inklusive ICC-profil-teknologien. Derfor, når man arbejder med en PDF, der er tiltænkt at arbejde i farvestyret workflow, så kræver det, at al farve bliver konverteret. Hvis konverteringen er til PS, så kan den originale ICC-profil ikke blive brugt i den senere konvertering - dog håndterer EPS en speciel instruktion (dvs. DSC-comment) – der muliggør ICC-profilens inkludering og senere brug. Dette virker imidlertid kun når profilen repræsenterer hele dokumentet og ikke hvis dokumentet indeholder flere forskellige profiler.

## Hvad skal man gøre?

Brug EPS og ikke PostScript!

Eftersom EPS er enhedsafhængigt, tilbyder support for at håndtere ICC profiler og som regel er nemmere at producere fra i nogle applikationer, så er det til enhver tid anbefalelsesværdigt, hvor refrying / dobbelt-konvertering ikke kan undgås, så at vælge at producere EPS i stedet for PostScript. Det er også at foretrække, at brugere selv foretager konverteringen, hvor indstillingerne kan kontrolleres, frem for at lade en applikation gøre dette for en, hvor man placerer pdf-filen som en billedfil. Ved at styre og kontrollere indstillinger, og sikre, at disse er valgt korrekt, er det muligt at holde et evt. databas til et minimum.

De følgende indstillinger er anbefalede:



- Eksporter som Level 3 PostScript, binært. Dette er den mest moderne og kompakte form for EPS, og sikrer dermed bedst muligt support for alle elementer (features) i den mindst mulige tænkelige fil.
- Inkluder alle indkapslede (embedded) skifter. Ved at inkludere alle originale skrift-data fra pdf-filen til EPS er der aldrig en risiko for, at Acrobat Distiller (eller en tilsvarende proces) re-placerer skriften med en tilsvarende.
- Tillad aldrig skrift-formater at blive konverteret, som fx TrueType til Type 1 eller en CID-kodet skrift til at blive en u-koded.
- Undlad at tillade farvekonvertering, med mindre det senere workflow kræver dette.

## Alternativer til dobbelt-konvertering

Mens refrying / dobbelt-konvertering af en PDF er uundgåeligt for at kunne integrere i et ikke-pdf-workflow – så er der ingen grund til at refry / dobbelt-konvertere en PDF for at modificere denne i et PDF-workflow. Der eksisterer mange værktøjer / applikationer både på server og på desktop-niveau, der muliggør en gen-kodning af PDF til en ny PDF, der bevarer de krævede pdf-karakteristika. Eksempler, som fx forberedelse for udgivelse på nettet, gennemføre farvekonvertering, skrifttype-korrektioner, filstørrelsesreduktion, er alle elementer, der kan gennemføres på den originale pdf-fil, hvilket altid vil resultere i en bedre pdf-fil til sidst i workflowet.

## Konklusion

Nogle gang kræver et specifikt workflow specifikt refrying / dobbelt-konvertering af en pdf-fil grundet brug af ældre (eller ikke pdf-egne) programmer / værktøjer. I så tilfælde er der ikke mulighed for at vælge og konvertering MÅ foretages. Imidlertid er der ikke nogen grund til at lade disse værktøjer 'drive' konverteringsprocessen. I stedet er det at anbefale, at manuel og specifik konvertering til EPS gennemføres, med brug af de ovenfor nævnte anbefalinger til indstillinger.

Mens man kan komme ud for nødvendigheden af at skulle konvertere en PDF til en EPS på grund af et ikke-PDF workflow, er der absolut ingen grund til at genere en Postscript fil for derefter at konvertere den til PDF.

Der er ingenting, der kan tilføjes via en refry / dobbelt-konverteringsproces, der ikke også (og bedre) kan tilføjes vil gen-skrivning af den originale pdf-fil. Eftersom refrying / dobbelt-konverteringsprocessen vil resultere i data-tab, bør en sådan ikke bruges.