



# Erfaringer som leverandør til bitarkivet

Søren Roug

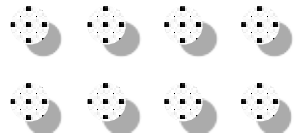
Charlotteskolen  
24 august 2023

# Min primære erfaringer (sådan gør jeg)

- PDF dokumenter
- Disketter
- Samlinger af disketter



Illustrations by Pixeltrue on [icons8](#)





# PDF dokumenter

- Skanner ren tekst i 200 dpi
- 300 dpi, hvis der er gråtone billeder
- Farvebrochurer kræver en afvejning. Der kan være billeder, som kan bruges til plancher senere.

Her skanner jeg med 600 dpi



# Vigtigt: OCR'et tekst

- Man kan søge i dokumentet
- Indekseret af Google – kan findes af andre
- Brug **ocrmypdf -l dan**
- Vigtigt med sprog – ellers genkendes æ,ø,å ikke!
- <https://ocrmypdf.readthedocs.io/>

# Document metadata

Nemmest at copy-paste fra OCR'et fil

- Document author: Hvis muligt. Kun personer – ikke firmaer
- Document date: Skal matche dokumentets dato format. Ville foretrække ISO 8601
  - Kunne være dejligt, hvis der var en dato-kolonne at sortere på
- Document description: Afsnit fra dokumentet, som beskriver indholdet. Produktnavne, andet ualmindeligt.

Brugeren skal vide hvad indholdet er, *inden* dokumentet åbnes.

# Disketter

- Bruger primært IMAGEDISK fra 3½" boot disk
  - Designet til at fremstille friske disketter.
  - Markerer dårlige sektorer. Nemt at korrigere med en hex-editor.
- Har en 486 PC med 5.25 HD drev
  - Læser både DD og HD disketter, men *ikke* 128-byte sektorer i FM format.
  - Bruger istedet en PC med 360KB drev i mikrorum #1
  - Også bedre til 40 spors disketter generelt

# Diskette metadata

- Media summary: Primære indhold inkl. version
  - Kan også skrive (platform, dansk/engelsk)
- Media type: 3½", 5¼" eller 8" Floppy Disk
- Media description: Mest muligt fra etiketten, Leverandør, Indholdsfortegnelse
  - CPMTools og Mtools

Brugeren skal vide hvad indholdet er, *inden* disketten bruges.

# Samling af disketter med BagIt

- Blev nødvendigt, da jeg havde indlæst 33 disketter til MS-Office
- Til alt hvor en distribution er på flere disketter

<https://github.com/LibraryOfCongress/bagit-python>

```
python -m bagit --contact-name 'Dansk Datahistorisk Forening' <dir>  
zip -r <dir>.zip <dir>  
edit <dir>.zip.meta
```



# Eksempler

## **Media.Summary:**

PolyDos 2.0 system master disk

## **Media.Description:**

SYSTEM MASTER

Serial number PD2210185

(C) 1983 Poly-Data microcenter

PolyDos 2.0 (G809/G815) Copyright (C)

1981 Piezodan Microcenter

PolyDos Disk BASIC Version 2.0.(C)

1981 J.Jensen microcenter

PolyZap Z-80 Disk Assembler Version

2.0 Copyright (C) 1981 J.Jensen

microcenter

## **Document.Title:**

A Short Guide to the Use of the CR8  
Microcomputer

## **Document.Date:**

Oktober 1982

## **Document.Description:**

Doc. reference:

CSD-MIC/001/USM/0009

This is intended to be merely the  
briefest of introductions to the CR8,  
for ...