

Specifications for File Formats

Martin R. Kalfatovic
**Smithsonian Institution Libraries &
Biodiversity Heritage Library**



**Workshop Coleção de Obras Raras
Essencial, São Paulo: 3-5 February 2010**

Digitization Specifications

- Minimum
 - Bitonal: 300 ppi
 - Grayscale :300 ppi
 - Color: 300 ppi
- Preferred
 - Bitonal: 600 ppi
 - Grayscale: 600 ppi
 - Color: 400 ppi

File Formats

- CR2: Camera Raw
- TIFF (uncompressed, LZW “lossless”)
- JPEG 2000 (uncompressed, “lossless”)
- JPEG 2000 (compressed, “lossy”)
- JPEG (very compressed, “lossy”)

Derivative File Formats

- PDF
- PDF B/W
- OCR TXT
- OCR with character coordinates
- Mobi (Kindle)
- Epub (e-readers)
- Daisy (audio readers)