Measuring Scanned Documents, Born-Digital Documents, & Digital Storage

Scanned Letter Size Pages
(All images are scanned 1 bit per pixel, black & white, and compressed, unless otherwise noted.)

1 scanned page (8 1/2 by 11 inches, A4) = 50 KiloBytes (KByte) (on average, black & white, CCITT G4 compressed)

1 file cabinet (4 drawer) (10,000 pages on average) = 500 MegaBytes (MByte) = 1 CD (Compact Disc) (ROM or WORM)

2 file cabinets = 10 cubic feet (cf) = 1,000 MegaBytes = 1 GigaByte (GByte);   8 file cabinets = 1 DVD-R (WORM) (see below)

2,000 file cabinets = 1,000 GigaBytes = 1 TeraByte (TByte);   2,000 file cabinets = 250 DVDs

1 box (in inches: 15 1/2 long x 12 wide x 10 deep) (400 x 300 x 250 mm) (2,500 pages) = 1 file drawer = 125 MegaBytes

1 box (packed) = 2 linear feet (500 mm) of files (loose enough for active filing) = 25 (rounded) linear inches = 125 MegaBytes

1 linear inch (~20 mm) = 100 pages = 5 MegaBytes;   1 thousand linear inches = 100 thousand pages = 5 GigaBytes

1 cubic foot (cf) (~.025 cubic meter) = 2000 pages = 100 MegaBytes; 10 cubic feet (~.25 cubic meters) = 20 thousand pages = 1 GigaByte

8 boxes = 16 linear feet = 2 file cabinets = 1 GigaByte; 8,000 boxes = 16,000 linear feet = 1,000 GigaBytes = 1 TeraByte

For paper and microform document imaging, see also AIIM (Association for Information and Image Management) [http://www.AIIM.org]

For records and information management, see also ARMA (Association of Records Managers and Administrators) [http://www.ARMA.org]

Scanned Engineering Drawings / Large Format Documents

1 E size drawing (48 inches by 36 inches) (A0 size) = 16 letter size pages (8 1/2 by 11 inches, metric A4) = 800 KiloBytes   To place an E size drawing in a file folder in a file cabinet drawer, the drawing must be folded in half 4 times and is then 16 sheets of paper thick when folded.

NB: Scanning must accommodate the older, untrimmed, US paper sizes, because it is the older drawings that are digitized by scanning.

Metric Trimmed Paper Sizes | United States Paper Sizes | Storage

Metric Name	Metric Size in Millimeters	Size in Inches	Number of Square Meters	Number of A4 Size Pages	US Name	New Size (Trimmed) in Inches	Old Size (Untrimmed) in Inches	Equivalent Letter Size Pages	Digital Image Storage Requirements
A10	26 x 37	1.02 x 1.46							
A9	37 x 52	1.46 x 2.07	1 / 512	1 / 32					
A8	52 x 74	2.05 x 2.91	1 / 256	1 / 16	Business Card	2 x 3 1/2			5 KiloBytes
A7	74 x 105	2.91 x 4.13	1 / 128	1 / 8	3 x 5	3 x 5			10 KiloBytes
A6	105 x 148	4.13 x 5.83	1 / 64	1 / 4	Microfiche	4.13 x 5.83			
A5	148 x 210	5.83 x 8.27	1 / 32	1 / 2	5 x 8	5 x 8			25 KiloBytes
A4	210 x 297	8.27 x 11.69	1 / 16	1	A	8 1 / 2 x 11	9 x 12	1	50 KiloBytes
A3	297 x 420	11.69 x 16.54	1 / 8	2	B	11 x 17	12 x 18	2	100 KiloBytes
A2	420 x 594	16.54 x 23.39	1 / 4	4	C	17 x 22	18 x 24	4	200 KiloBytes
A1	594 x 841	23.39 x 33.11	1 / 2	8	D	22 x 34	24 x 36	8	400 KiloBytes
A0	841 x 1189	33.11 x 46.81	1	16	E	34 x 44	36 x 48	16	800 KiloBytes
2A0	1189 x 1682	46.81 x 66.22	2	32					1.6 MegaBytes
4A0	1682 x 2378	66.22 x 93.62	4	64					3.2 MegaBytes
					F	28 x 40	varies		600 KiloBytes
					G	11 x (22 1 / 2 to 90)	varies		
					H	28 x (44 to 143)	varies		
					J	34 x (55 to 176)	varies		
					K	40 x (55 to 143)	varies		

Paper size references: US MIL-M9868-D, Microfilming of Engineering Documents, 35MM, Requirements for, 10-1-70 and amendments 1 and 2, 2-12-82 and 9-20-82; MIL-STD-804B Format and Coding of Aperture, Copy and Tabulating Cards Engineering Data Micro-reproduction System, 15 August, 1966; ANSI Y 14.1 [http://www.ANSI.org], 1980, Drawing Sheet Size and Format, published by ASME (American Society of Manufacturing Engineers) [http://www.asme.org], New York; Metric standards first published in 1922 by DIN (Deutsches Institut für Normung) (German Institute for Standards) [http://www.DIN.de] Now used worldwide as ISO 216 [http://www.ISO.ch]. (See page 8 for ISO references)

Sizes G, H, J, and K are US roll sizes. Letter Size is A4, A, or 8 1/2 x 11; depending on the system of measure.
Ledger Size is 11 x 17 inches.
Tabloid Size is 11 x 17 inches.
Legal Size is 8 1/2 x 14 inches.

Newspapers: A double truck (center fold) full broadsheet is 24 in x 36 in, equivalent to an old D size drawing in size. Because a newspaper page would be scanned at a higher resolution and contains detailed graphics, a double truck would require about 1 MegaByte and a single full broadsheet page (18 by 24 inches) would require about 1 / 2 MegaByte. See also NAA (Newspaper Association of America) [http://www.NAA.org]

Scanned Microforms
1 roll of 16 mm microfilm (100 ft, ~30 meters) (24X reduction) = 2,500 letter size images = 1 box = 1 file cabinet drawer = 125 MegaBytes
1 roll of 35 mm microfilm (100 ft) (12X reduction, open spacing, normal scan) = 1,000 letter size images = 50 MegaBytes

1 microfiche (105 mm film) (24X reduction) = 100 letter size images = 5 MegaBytes (average); 200 microfiche = 20,000 images = 1 GigaByte

In many record series, each microfiche contains only a few images because each fiche represents a single record in the series (e.g. one fiche per person in a personnel record series). In this case filming breaks on records, rather than being continuous. To a lesser extent this is also true for roll film. In these cases, the amount of storage required depends on the number of images on the film, not the number of microfiche or the number of rolls of film. A full, standard 24X microfiche has 7 rows of 14 letter size (8 1/2 x 11 or A4) images for a total of 98 images.

As with any microform, scanned aperture card images require the same storage as images scanned from the paper original of the document in the aperture.

Compression

All documents are stored and transmitted in compressed format. All compression formats are assumed to be lossless or used with a lossless setting, except MPEG (Moving Picture Experts Group), unless otherwise stated. Lossless or non-destructive compression (as opposed to lossy or destructive compression) does not change the document. That is, a decompressed document is identical to the original document before compression was done. Lossless compression is often needed to meet legal requirements for document storage. The most common form of one bit (per pixel), bitonal (The two tones of color are two shades of grey which are black and white.), lossless compression, used in TIFF G4 and Adobe PDF (Portable Document Format), is the CCITT G4 (Group 4) facsimile compression format. Before using any other form of compression, it is often useful to evaluate the cost savings of moving to the less common format. The CCITT (Comité Consultatif International pour le Télégraphe et le Téléphone) (International Telegraph and Telephone Consultative Committee) is now a part of the ITU (International Telecommunications Union) [http://www.ITU.int]   The G4 ITU recommendation T.6 (11/88), Facsimile coding schemes and coding control functions for Group 4 facsimile apparatus, is on pages 48-57 of the CCITT Blue Book, Volume VII - Fascicle VII.3, Terminal Equipment and Protocols for Telematic Services, Recommendations T.0 - T.63, ISBN 92-61-03611-2

Size of Born-Digital Documents: PDF, COLD, Word Processor Pages, & OCR Output Pages || Commercial Scanning Cost

PDF (Adobe.com Portable Document Format) files, created by scanning and lossless compression, are about the same size as other compressed raster scanned documents. PDF files (vector files) created directly from word processor or print files are about 10 KiloBytes per page for text pages, 50 KiloBytes for graphics (line art), and about 1 MegaByte for pages containing mostly images. High resolution (1200 dpi 24-bit color) letter size images can be as large as 250 MegaBytes).

Microsoft Office 2000 (and later) documents: 20 KiloBytes per document for the XML repository + 5 KiloBytes per page //// 1 wordprocessor page, 1 office suite page (without the Office 2000 XML repository), or 1 OCRed (Optical Character Recognition) page = 5 KBytes (all pages listed above this section are scanned pages) For SGML (Structured Generalized Markup Language), HTML (HyperText ML), XML (Extensible ML), and CGM (Computer Graphics Metafile): see OASIS (Organization for the Advancement of Structured Information Standards) [http://www.OASIS-open.org] W3C (World Wide Web Consortium) [http://www.w3.org/XML]

1 compressed page of COLD (Computer Output to Laser Disc) or COOL (Computer Output On-Line) (including index) = 2 KiloBytes for letter size statements, 4 KiloBytes for 11 x 14 inch fanfolded greenbar computer sheet, 10 KiloBytes for All Points Addressable (APA) pages such as IBM AFP (Advanced Function Printing) and Xerox Metacode. For printing, see also Xplor International [http://www.Xplor.org]

Minimum commercial scanning cost for backfile conversion (more than 1 million pages) = about ~ 5 US cents per page

Compiled by Steve Gilheany, CRM, CDIA   Trademarks are the property of their respective holders.