

What's Your Measurements? Version 2.0

I call this Version 2.0 like the software companies do when they release updated versions of their products, as I originally wrote a similar article in 2002. At that time, not much was known about the relation of electronic documents to page counts. Since then, the processing of millions of pages of Ediscovery (electronic documents) has provided sufficient data to make reliable estimates based upon those experiences. Of course, we're talking about how to determine the number of pages for an electronic collection. But first, let's review and update paper collection estimates.

PAPER COLLECTIONS

THE GUESSTIMATING METHOD

The following is typical for guesstimating paper collections where the box or drawer is full:

Legal/Letter Box	2,500 pages
Long Bankers/Transfer File Box	4,500 pages
Vertical File Drawer 18"	3,600 pages
Vertical File Drawer 24"	4,500 pages
Lateral File Drawer 36"	6,000 pages
Lateral File Drawer 48"	8,000 pages

How many times has your attorney asked, "How many pages do we have?" How many times have you asked a vendor, "What will it cost and how long will it take?" How many times have you heard "IT DEPENDS?" It depends largely upon the total estimated number of pages and the deadline when you need it completed. The rest is a matter of throughput calculations based on the number of machines and personnel available to accomplish the task.

Below is another method you can use to estimate the size of your paper collections. It is based on pages per linear inch. Take your measurements by putting your hand behind the files and gently pressing them forward, then measure the number of inches of paper and use the multiplier below.

THE PAPER MEASUREMENT MULTIPLIER

Type of Paper	Pages per Inch
Photocopies, no bindings	248
Photocopies, with binding elements	225
Originals, w/bindings and folders	200

A ream of standard 20-lb paper contains 500 sheets and measures 2 inches for 250 pages per inch. Of course, the paper is tightly compressed and contains no printing, binding elements or folders. Printing adds very little expansion, while binding elements and folders do add thickness resulting in the lower multiplier used.

In my 2002 article I wrote that there is no way to even guess at how many pages are involved with e-discovery electronic files! Much has been learned from experience in handling millions of pages since that time, providing some indications that may be used to estimate electronic document collections.

"E-DISCOVERY" - ELECTRONIC DOCUMENT COLLECTIONS

HOW DO YOU ESTIMATE IT?

- 1 Megabyte (MB) will average around 75 pages
- 1 Gigabyte (GB) will average around 75,000 pages
- 1 Terabyte (TB) will average around 75,000,000 pages

- Email average 1-2 pages per each
- Word Processing File average 5-8 pages per each
- Spreadsheets average 15-30 pages per each
- Presentation average 12-24 pages per each
- Graphic average 1 page per each
- Adobe PDF File average 35 pages per each

- Diskette 1.44 MB if full 50-150 pgs
- Zip Disk 100 MB or 250 MB if full 7,500-18,750 pgs
- CD 640 MB - 800 MB if full 48,000-64,000 pgs
- DVD 4.7 GB - 17 GB if full 350,000 - 1.3 million
- Tape Drive 2 GB - 360 GB if full 150,000 - 27 million
- Hard Drive 20 GB and over if full 1.5 million and up

Now, please, this is not an exact science and someone will surely measure or determine his or her collection to find a different number. Remember, we're talking estimates here.

Electronic files are received without regard to order or type. The vendor should review the directory listings to determine the number and sizes of the various file types. Once a file list analysis is performed the attorney can utilize the information to determine which files to review before converting to image or paper, providing greater control over the cost involved in processing e-discovery. Applications or system files may be of little value and not worth processing.

Good estimates help to avoid problems when the vendor provides discounts for size and collection is substantially smaller or when the collection is substantially larger than estimated and your attorney must go back to the client for additional funding as a result.

Open Door Solutions, LLP is a Dallas-based company providing litigation support services and document management solutions to law firms and corporations nationwide. Author Bob Sweat received his education in Business Administration and Economics at the University of Wisconsin and advanced work at Purdue University. He holds a Paralegal Certificate in Civil Litigation with Computer Emphasis from The Center for Legal Technology, Milwaukee, WI, and has years of experience working with local and national vendors on large, complex litigations. Bob is currently a partner at Open Door Solutions.