Digital Document Specialist – Level I

NasonProductions - Educational Series

original document:

08/22/2006 by Scott Nason

Every effort has been made to ensure the accuracy and integrity of this material. This material is presented “as is” and there is no warranty expressed or implied as to its correctness.

Table of Contents

DDS Overview

What is a Digital Document Specialist

A **Digital Document Specialist** is an individual that is **comfortable** engaging in the process of **creating** and **distributing**, **storing**, **retrieving** and **managing** digital or electronic documents.

> *One who **specializes** in dealing with document-based **digital information**.*

Intended Audience

This class is a **technical** overview of the **use and creation of digital documents** and is **intended** for individuals that, **through the normal course or their daily activities**, would be involved in the **manipulation**, **production** and **distribution** of **electronic documents**.

This could include **DMS employees** involved with a customer's **digital workflow** or **in-house personnel** involved in almost any **digital document process**.

Objectives

This course will provide knowledge of the **processes and technologies surrounding digital documents** from **creating digital documents** to **scanning** image files into digital form.

After successful completion of this course, the participant will **understand of how to deal with various types of digital documents** within common **digital workflows**.

Labs in **Adobe Acrobat workflow** are coupled with both **PUSH and PULL** scanning **hands on exercises**.

The class is **approximately 60% theory and 40% hand on**. At the end of the day, there will be a test and a **certificate of completion** for those achieving a passing score.

THIS CLASS IS INTERACTIVE: ASK A LOT OF QUESTIONS!

Prerequisites

General **computer literacy** including an **understanding of Windows** and a familiarity with the **use of a PC** is necessary.

Class Outline & Timetable

08:00 – 08:30	Introductions, expectations, day layout & general housekeeping
08:30 – 09:00	Digital Documents Introduction
09:00 – 10:15	PC Literacy Primer
10:15 – 10:30	break
10:30 – 12:00	File Formats, Storage & Retrieval, Electronic Documents Lab
12:00 – 01:30	Lunch
01:30 – 03:00	Scanning
03:00 – 03:15	break
03:15 – 04:30	Acrobat Lab
04:30 – 05:00	Final Exam

Digital Documents Introduction

What are Documents?

doc-u-ment (doky-mnt)n. *Abbr.* **doc. 1.** A written or printed paper that bears the original, official, or legal form of something and can be used to furnish decisive evidence or information. [1]

*A document is a **container** for **information**.*

For our purposes, we're going to define documents as **containers for information**. A piece of paper is a container of information. A bill, an invoice, a letter, a contract; these are all various kinds of information. We use documents to **distribute**, **share** and **save** information. A **filing cabinet** is a **database** of structured information.

Paper has been, and will continue to be, an excellent method for the **storage and sharing** of **information**. The technology surrounding paper and paper-based information has been around for many years and is **well established**. People are **comfortable** with the processes surrounding a paper-based system and generally can navigate through most any system with little or no training.

But paper has its **shortcomings**. Paper **wears out**; it will **discolor** and **deteriorate** with age. Paper can be **damaged** easily with heat, open flame or liquids. Paper has to be copied in order to be shared among multiple users simultaneously. This can create **versioning** conflicts and questions about which documents to rely on.

What's Digital?

If we're going to talk about **digital documents**, we probably have to know exactly **what we mean** by digital.

Digit; to represent with **numbers**; verses analog – to represent with a smooth continuous change. In computer-land that means 1's and 0's; on or off.

*The **Binary** number system uses only zeros and ones.*

Explain the use of the word digital in these;

- digital computer
- digital camera
- digital thermometer
- digital cable TV

Digital Documents

[1] Excerpted from *American Heritage Talking Dictionary*. Copyright © 1997 The Learning Company, Inc. All Rights Reserved.

Probably going to be interchanging **electronic** and **digital** a lot (means same for the purpose of this class).

When we talk about **digital documents**, we're talking about a broad range of **information** that is currently **stored** in **electronic (digital) format**. These can be documents that **originated on the computer** (Word, Xcel, etc.) or they can be documents that were **converted** to electronic format like **scanned image** files.

*A document is a **container for information** and there are **two main types** of digital documents –*

***electronic** generated*
and
***converted** (scanned).*

Digital Document History

To help us get a sense of where we are today with digital documents, let's take a short look at where we've come from. The entire history of digital documents doesn't go that far back.

Before computers came along, **desktop publishing** utilized the original **cut & paste**; with actual **scissors** and **glue**.

The Desktop Revolution

2 things were responsible for the desktop revolution; **Macintosh** & the **Laser Printer**

- 1984 - The **Apple Macintosh** debuts.
- 1984 - Hewlett-Packard introduces the LaserJet (Canon's A-1 engine), the **first desktop laser printer**. (what was available at this point in time? **Dot matrix** and **daisy wheel**; professionals used **typesetters**)

Third component comes along:

- 1985 - **Adobe** introduces **PostScript**, the industry standard Page Description Language (PDL) for professional typesetting.
- 1985 - Aldus develops **PageMaker** for the **Macintosh**, the first "**desktop publishing**" application. CUT & PASTE ON THE COMPUTER.
- 1985 - Apple produces the **LaserWriter** (Canon's A-1 engine), the first desktop laser printer to contain **PostScript**. – **now the stage is set.**
- 1987 - PageMaker for the **Windows** platform is introduced.

The **Macintosh** was originally considered the **only computer** to perform any **high-end graphic** applications on. Today, with the advent of **Windows XP** (and soon to be **Vista**), the gap has narrowed (if not all but disappeared – the debate still rages on).

Digital Documents Today

Today what factors **contribute to the use** of **electronic (digital) documents**?

- Availability of technology (scanners, MFPs, inkjet printers, digital cameras)
- Acrobat & PDF

What **electronic documents** do you use in your daily lives?

- Email
- MS Office (Word, Excel, Powerpoint)
- The Web
- Digital Photos
- e-books

Digital documents are **almost everywhere** today. They are responsible for an **information explosion**. We have access to **more information** than **any other generation** in history.

PC Literacy Primer

Assessment

This class is designed around the PC and does not cover any use of the Macintosh. Using a **PC** is an **essential part** of dealing with digital documents. If you're **uncomfortable** with the **basics of a computer**, then you will be **more than uncomfortable** with the basics of digital documents.

On a scale of **1 to 10**; how would you **rate yourself** on a PC (need to be at least a 4)

We're going to go over some **real basics**. May be review for a lot of you, but if I use any word or **terms you don't know**, **please stop me and ask**.

> *Understanding and being **comfortable** with a PC is an **absolute essential ingredient** to mastering electronic documents.*

What is a Computer?

A **computer** basically has just **3 parts**. Think of it like a car; **engine**; **storage** & **interface**.

The Car:

- **Engine**; provides all the horsepower for getting the job done; moving from point A to point B.
- **Storage**; the trunk; where you keep the stuff that you need to have with you on your trip.
- **Interface**; the part that the end-user (driver) interfaces with including the steering wheel, brakes, gear shift and all the dashboard controls.

The Computer:

- **Engine**; the processor or CPU; what does all the work or computing.
- **Storage**; where the information is saved; the hard drive. (or temporarily in RAM) volatile verses non-volatile.
- **Interface**; the part that the end-user interfaces with including the **mouse**, **keyboard** and all the **windows** on that you open.

Storage

> *"But I saved it in **Word** and now I can't find it."*

Understanding **how your computer stores information** is probably the **single most important** key to mastering your comfort level with a PC.

Information called **files** is placed—in **digital** form—on **physical devices** like hard drives, thumb drives and diskettes. This information can be grouped together in special file containers called **folders**.

There are **three** main ingredients to **finding any file** on your computer:

- DEVICE
- FOLDER(S)
- FILE (NAME)

My Computer

The part of a computer that we're going to have the **most interaction** with is the **interface** and the **storage**. If you can't **effectively navigate** the Windows storage system then you're **dead in the water**.

My Computer is your window into the storage on your computer.

Understanding **My Computer** (and its refined baby brother; **My Documents**) is one of the **most important** things that you can do to **ensure your mastery** of the PC. **My Computer** is what we use to **see** what's in our computer. It's one of the **principal ways that we find and interact** with files (information).

Essential My Computer knowledge:

- Moving up and down 1 directory
  - Inside of a folder is "drilling down"
  - My Computer is the "root" of the tree
- Address window; informational & quick navigation
- The "Back" button is different than "Up"
- View: Thumbnails, Tiles, Icons, List, Details
- Turn on file extensions display
- Make all folders the same

Viewing File Extensions

The way that Windows is **configured** will determine whether or not you can **see the file's last name** or **extension**. Windows **default behavior** is to **hide the file extension**—it's its way of **protecting** you from yourself. It's also a way for a **PC to be more MAC-like** which never had file extensions.

If you want to be able to see the file's last name in Windows: in **My Computer** select:

- Set your current **My Computer** View to **Details.**
- **Tools | Folder Options | View** and then uncheck the box that says **Hide extensions of known file types.**
- **Click** on the **Apply to All Folders** button.

Let's talk about the some of the **terms** that have to do with **My Computer** and the **storage devices** on your computer.

Device Names

The computer assigns a **letter** to a **physical storage location or device** – it names it so that **we have a way** of referring to it. The **colon** after the letter tells the computer that it's a **logical device name** and **not a file name**.

- A: is different than A

*"A:" represents a **physical** storage device on the computer while "A" represents a **file name** – **this is important**.*

The **colon** says, "*I'm a physical device or thing.*" For that reason, a **colon is an invalid character** to use in a **file name**.

Physical storage devices are the actual **pieces of hardware** (or devices) attached to your computer (drive letters are **examples only** and may actually be different):

- Hard drive c:
- Floppy diskette drive a:
- A network drive i:
- CD or DVD drive d:
- Zip drive f:
- Tape drive r:
- USB drive g:
- Memory Stick Reader h:

*Windows **is not case sensitive** for file names, folders, devices or user names.*

Physical verses Logical

A **logical device name** is a map or a **pointer** to, either a **physical device**, or a *portion* of a physical device. For example, **A:** refers to a **physical floppy diskette** drive while **I:** may refer to a **mapped network drive**.

I: refers to a **logical device** name, but as far as Windows (and you) are concerned, it's a **physical device**.

Drive letters **A:** and **B:** are reserved for floppy diskette drives and **C:** is reserved for the first hard drive on the system. All of the other letters may be assigned to other **physical devices**, **hard drive partitions**, **network drives**, etc.

What Is A File?

From your computer's standpoint, what do all of these things have in common?

- folders
- programs
- directories
- sub-directories
- documents
- pictures
- videos
- sounds
- music
- spreadsheets
- databases

They are all *files*. And **all of the files** on a computer fall into **four** categories:

- PROGRAMS (or EXECUTABLE)
- DIRECTORIES (or FOLDERS)
- SHORTCUTS (or POINTERS)
- DATA

To understand how information is stored inside of our computer, we need to understand how the computer sees the **stuff** that we put into it. A computer's first and **most important task is to deal with files**; to **organize** them and to **execute** them; to **copy** and **store** them. But what are these files and what can we do with them?

After we talk about how to **name a file**, we'll talk more about the **4 kinds of files**.

File Naming

In order for files to have any use to us, we first have to have something to call them. Now that we know that there are only four kinds of files, what (and why) do we call the individual files? We need a little history lesson here so that we can understand the present.

In the beginning (DOS 1.0), files had a **first name** and a **last name**. First names could be any combination of letters and numbers and could be a **maximum of eight characters long**. The last name could also be any combination of letters and numbers, but they were only **three characters long** – and they were **optional**. The first and the last names were **separated by a period** so you'd end up with something that looked like this:

- myfile.doc
- simple.txt
- letter1.ltr

Today, with **Windows**, we can have file names that have spaces and are around **250 characters long**. In practice, it's better to keep file names **shorter** and **to the point**; so that they are **easily identifiable** as to what they are.

There are a lot of **rules** to naming files but here's a **general outline**:

- Try not to use file names over **50 characters** or so.
- Don't use **special symbols** other than the underscore or dash.
- Don't rename the **file extension** to something other than expected.
- File names in Windows **are not case sensitive** (not true for other platforms).

Remember, just like your name, a file name is how that piece of information will be identified later.

The File Extension

The **file extension** or last name of a file is used for **two** main purposes:

1) To **register** itself with Windows (the operating system).
2) **Human beings** can use the file extension to know **which program created** the file.

The last name ('**doc**' in our 1st example) is referred to as the **file's extension**. **Sometimes**, the file extension tells us something about **what kind of file it is**. But, it's important to remember that the name as well as the extension can be any combination of letters and numbers and therefore **could have no meaning at all** – except to the person that created it.

Windows file names are **case insensitive**. That means that **unlike** your Windows or network **passwords**, filenames **don't care about upper and lower case**.

- **my file name.doc** is the same as **MY FILE NAME.DOC**

With Windows, there is **no stipulation about the length** of (or **even having one at all**) the last name of **file extension**. This is a legal file name in Windows:

- my file name.document from Word

For the most part, though, programs still **adhere to the 3 character file extension for backward compatibility**. And it's generally **not a good practice to change** the file extension if it already exists.

File Registration

*"I just got an email with an attachment that **I can't open**."*

When an application program (Word, Excel, Acrobat, etc.) is installed, it **registers** these last names with the **Operating System** (Windows) so that it knows what programs are associated with what **file extensions (last names)**.

When you **double click** on a **program file**, the program **runs** or executes. It starts up its user interface and presents you with the options available in that program. When you double click on a data file, something else takes place.

If the last name of the data file is **registered** in Windows, then Windows is aware of what **program created this file** and will **run it** automatically. If the file extension is not registered, Windows doesn't know which application created this file **so it asks you the question**.

If you know that a particular document was created in **Microsoft Word**, for example, you could scroll down the list to **Microsoft Word** and then click on OK to open up the data file using the application program, Microsoft Word.

If you know that this is a file extension that you'll probably run across again, you can check the box that says "**Always use this program to open these files**". Checking this box registers the file extension (or last name) with Windows. Now, any time you click on a file with this newly registered file extension, it will open up with the program that it's registered to (Microsoft Word in this example).

If you do not want to register a file extension with a particular program, you can always **Start** the program first and then use **File | Open** to get the file opened in the program.

Some **common** file **registrations** (or file associations) are:

- DOC >> Microsoft Word
- PDF >> Adobe Acrobat
- XLS >> Microsoft Excel
- PPT >> Microsoft PowerPoint
- PSD >> Adobe Photoshop
- QXD >> Quark Xpress

Here are some common file extensions and their **expected** meanings. We'll look into some of these further as we explore the different file types.

• Executable files	EXE
• Document files	DOC
• Portable Document Format	PDF
• A Window's System file	SYS
• Text files	TXT
• Help files	HLP
• Overlay files	OVL
• Adobe Photoshop	PSD
• Adobe Illustrator	AI
• JPEG Picture file	JPG
• TIFF Graphic Image file	TIF
• Database file	DBF
• Window's Dynamic Link Library	DLL

Now let's look a closer look at the **four different kinds** of files.

The 4 Types of Files

Program Files

Program files are special files also known as **application files**. They are the only files that can **do something**. Program files can be **run** (or **executed**). All other files are either, **used or acted on, by program files**. Program files are the ones doing the work. The Operating Systems that we use (Windows 98, W2000, NT, XP) only recognizes 4 last names (or **file extensions**) for these **executable files**.

• Executable files	EXE
• Small executable files	COM
• Batch files	BAT
• 32 bit Command File	CMD

The files that end in **EXE** are by far the **most common executable files** on any computer today. These **program files** are designed to do specific tasks like **word processing**, **spreadsheets** or **play CDs**.

Directories

The terms **directory, subdirectory, folder** and **sub-folder** all mean the same thing. We're going to end up using these terms interchangeably so just remember they're all the same. They're all a special kind of file whose only job is to act as a **storage location** for other files.

Directories (or folders) are just **holding areas** that allow us to **group other files together**. Folders and directories have the **same naming conventions** we use for all other files. Although it **is valid to assign a last name** (or a file extension) to a folder, in practice this may be confusing.

Directories can be **deleted**, **moved** and **copied** just **like other files**. The only thing is, **when you delete a folder, you delete all the files in that folder**. When you move a folder, you move all the files in that folder along with it.

It's **perfectly acceptable** (and advisable) to **have folders inside of other folders**. This simplifies and **clarifies the organization** of the information on your hard drive.

Folders are going to be one of our **main tools when it comes to organizing** our computers and making them **work for us**.

- Right-click on your Desktop and select **New | Folder**
- A new folder will appear on the Desktop called **New Folder**
- Rename the folder to **DDS**

Shortcuts (or Pointers)

Shortcuts are **pointers** to other files. That means that when you execute (or open) a shortcut, you're actually executing (or opening) **the file that the shortcut points to**. When you **delete**, **copy** or **move** a shortcut, you **haven't deleted copied or moved the original file**. You've **only deleted** (or copied or moved) **the pointer** to that file or directory.

- Right-click & hold on **DDS** and drag it over a little
- When you let go of the mouse button, select **Create Shortcut Here**

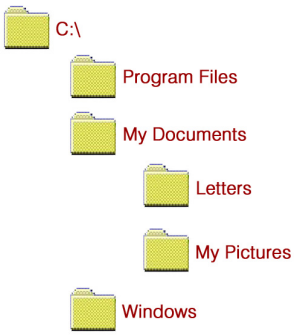
Data Files (or everything else)

Data files make up the **majority of all files on your computer**. One program file may require **hundreds of other files** to run correctly. There are data files that are actually extensions of programs (.DLL) and there are data files that are only used by the Operating System (.SYS) and there are data files that are used to store specific information about how programs startup and operate (.INI).

There are literally **thousands of data file types** and it really is impossible to know what they all do. So let's look at some of the most common one's that you'll run across.

- DOC Microsoft Word document
- PPT Microsoft PowerPoint Presentation
- BMP Windows bitmap raster image file
- AVI video file
- GIF web graphic file
- XLS Microsoft Excel workbook
- WAV an audio wave file
- TIF a raster image file
- JPG web graphic file
- MOV Quiktime movie file
- INI an INItialization file

Directory Structure



Windows uses what's called a **Tree Structured Directory** to organize if information. The physical drive letter is referred to as the **root** and the folders **branch out** from there.

Within this directory structure, files have a **complete pathname** – or a unique way of finding them on the tree. The pathname always begins with the **drive letter**, followed by a **backslash** character. For this reason, **we can have the same file name in multiple locations** on our hard drive but they will each be **separate files**.

- C:\My Documents\Letters\My New Letter.doc
- D:\My Documents\Letters\My New Letter.doc
- C:\My Documents\Letters\personal\My New Letter.doc

The **first backslash character in a filename refers to the root** of the drive. **C:** means "**in the root of the physical C: drive**". From there we add one or more **folders** (sub-folders) **separated by backslashes**, ending with the **file name** we are referring to.

If we changed the letter **C** at the beginning of the pathname to **D**, then we'd be referring to a completely different filename.

Files Sizes & Computer Math

Digital documents take up **space** on our computer storage. We need to be able to talk intelligently about **how big a file is**. A lot of really big files could contribute to us **running out of hard drive space**, bogging down an email server or giving a your network administrator an ulcer.

File Size: The Basic Units of Measure

- byte – computer lingo for one character or one letter
- kilobyte – one **thousand** (**three** decimal places: 000)
- megabyte – one **million** (**six** decimal places: 000,000)
- gigabyte – one **billion** (**nine** decimal places: 000,000,000)
- terabyte – one **trillion** (**twelve** decimal places: 000,000,000,000)

It's important to understand the **relationship** between these numbers—and you can do it easily by **adding or taking away zeros**.

1,000 kilobytes is a **megabyte**: 1,000 (3 zeros) plus kilo (3 zeros) = 6 zeros or mega.

1,000 megabytes is a **gigabyte**: 1,000 (3 zeros) plus mega (6 zeros) = 9 zeros or giga.

Remember Where the Zeros Are

- 1.6 megabytes = 1,600,000 bytes
- 1.6 gigabytes = 1,600,000,000 bytes
- 1.6 kilobytes = 1,600 bytes
- .6 kilobytes = 600 bytes

A Little Math

Storage capacity is measured in bytes. A 3 ½ inch floppy diskette can store 1.4 **megabytes** of information. One megabyte equals one thousand kilobytes. So our floppy diskette can hold (add six decimal places) 1,400,000 characters of information. Or, to say it another way, 1,400 kilobytes.

Lets say that a typical Word document takes up **80 kilobytes** (80,000 characters or **80k**) of space. How many documents of this size with fit on our 1.4M floppy disk?

- 1,400,000 / 80,000 = 17.5

So, if you've got more than 17 documents you need to copy to a floppy diskette . . . well, you get the picture.

Hard drives are rated in gigabytes (nine decimal places - 1,000,000,000). A **gigabyte is one thousand megabytes**. Typical hard drives range in size from around 60 gigabytes to over 500 gigabytes.

How many of our **80k** documents will fit on a **60G** hard drive?

- 60,000,000,000 / 80,000 = 750,000

The User Interface (OS)

OS or **Operating System** – is there for what purpose? It's **for people to interact with** (since we can't talk "computer"). Just like our car analogy, the OS gives us a way of **interfacing with the technology**.

Windows is an **Operating System** and in controls all **human interaction** with the computer through devices like the **mouse**, **keyboard** and the behavior of the **windows on the monitor**.

The Mouse

In a Windows-based OS, the **mouse is probably as important as the keyboard**. It may seem trivial, but a **solid understanding** of the how to **use the mouse** is essential to your overall comfort level with a PC.

Using the Mouse

Pressing a mouse button is called **clicking**. Pressing a button twice quickly is called **double-clicking**. The motion of the mouse controls the motion on the screen of the **pointer** or **cursor**. The **shape of the pointer or cursor** changes depending on where it is and what is happening. The term **cursor** is used when the shape shows where your typing will appear. Otherwise the term **pointer** is a better choice.

Some common mouse pointers-cursors:

-  standard pointer
-  waiting for something to occur
-  text mode
-  link or button
-  move something

*Using your mouse **comfortably** takes practice. This is a physical skill that uses muscles that you **may not have used** in this way before.*

If you are told to **click** on something, you should move the pointer over the object, press the **left** mouse button, and then release it.

Drag means to hold a **mouse button down** while **moving the mouse**. Usually **something** on the screen will either **move** or **become highlighted** as you do this.

Anatomy of a Click

> *"If I don't say **right-click**, then it's a **left-click**"*
> *- Anonymous*

There Are 5 Mouse Clicks:

- Single (generic mouse click)
- Double
- Right
- Drag
- Right Click n Drag

Single

Single-clicking on an icon or file is always done with the **left mouse button**. The generic term **click** means **left-single-click**. A left-click is most commonly used for **selecting** or **highlighting** a file or icon. Often a left-click will **initiate** something like a **link** or a **button**.

NOTE: When performing an action with a single left-click, the action usually **does not occur until you release the mouse button**. Like a chess move that isn't official until **you let go of the piece**, you can hold the mouse button down for a moment and as you make your final decision about continuing or not. If you decide not to perform the action, **instead of releasing** the mouse button, just **drag it off of the button** or item first.

Right-Click

Right-clicking on an item brings up a **Context Menu** for that item. The items on this menu change **depending on what you right-click on**, and they usually are a list of actions that pertain to that item. This **shortcut menu** provides convenient access to the various functions you can perform on the item.

For example, right-clicking on the **Desktop** brings up a **context menu** with settings **specific** to the **Desktop**. Right-clicking on the **Taskbar** brings up its own context menu. Most context menus have an entry called **Properties.** This gives you **specific settings** and **information** pertaining to that object.

- Right-Click on the **Desktop** – look over the various selections then select **Properties**.
- Right-Click on the **Taskbar** – under **Toolbars**, turn off **Quick Launch** (then back on)

Double-Click

Double-clicking always refers to **left-clicking**. There are **no functions in Windows** that use a **double right-click**. A double-click is **two short left-clicks within a very short period of time**. The interval that the computer will recognize as a **double-click** is **adjustable in the Control Panel**.

Double-clicking an **icon** or **file** will usually **execute**, **run** or **open** that file. In most programs, **double-clicking a word will select** (highlight) that one word. Double-clicking on the **Title bar** of a window will **toggle the full-screen mode**.

Drag

By clicking and **holding down the left mouse button** down, you can **grab an object** like a **file** or a **folder**. While holding the left mouse button down, **move the mouse** and the file or folder will move as well; in effect **dragging** it along with the mouse.

Right-Click n Drag

A **right-click n drag** has a very **specific behavior** in Windows which is **very powerful**.

When you **hold down the right mouse button** and then **drag** an icon or file, Windows displays a **context menu** that gives you the option to either, **move** the object to the new destination, **copy** the object to the new destination, **or create a shortcut** to that object and leave the shortcut in the new destination.

*Right-clicking to **move** or **copy** is a **great habit** to get into because **you** are **in control** of what Windows does.*

Drop

One more term used in and around mouse etiquette is **drop**. Drop means to **let go of the mouse button**. When you click and hold on an object like a file or folder, you can move that file or folder with the mouse. When you **let go of the mouse button**, you let go of the item—**dropping** it at the intended location.

Selecting Files

- Single-click
- Shift-click
- Ctrl-click
- Lasso
- Ctrl-A

Single-clicking on a file or object **selects** or **highlights** that file or object. **Selected objects** are displayed in **inverse-video**. You can **add to the list** of selected objects by holding down the **Control key** while **clicking on another selection**.

Alternatively, if you want to select a **group** of **adjacent icons**, you can click on a blank area – **not on the icon or file**, and **drag a box** around those icons or files with the mouse.

Holding down the **Shift key** while clicking on something will highlight (or select) all the icons or files **between two selections**.

Control-A (*hold down the* **Ctrl** *and* **A** *keys together*) will highlight all items in a folder or directory.

- Open up **My Computer** and **open** (double-click) the **D: drive.**
- Highlight
- Ctrl-Click
- Shift-Click
- Ctrl-A
- **Copy** each of the folders to your **DDS folder** on the Desktop using the **Right-Click n Drag** method.

Keyboard Shortcuts

Keeping your hands on the keyboard when they're already there can **increase productivity** by allowing you to get things done **more quickly**. If you're constantly moving your hand from the keyboard to the mouse that can **waist time** and lead to an **uncomfortable working experience**.

Windows has a lot of **keyboard shortcuts** built in. These are **keyboard equivalents** to things that you'd **normally do with the mouse**. Learning just a few of them can help speed up your work tremendously.

Windows **program key** combinations:

- **CTRL+C : Copy**
- **CTRL+X : Cut**
- **CTRL+V : Paste**
- **CTRL+Z : Undo**
- **CTRL+A : Select All**
- CTRL+B : Bold
- CTRL+U : Underline
- CTRL+I : Italic

Windows **system key** combinations:

- F1 : Help
- CTRL+ESC : Open Start menu
- ALT+TAB : Switch between open programs
- **ALT+F4 : Quit program / Close the active Window**
- SHIFT+DELETE : Delete item permanently
- **ALT+any underlined menu item**

Shell objects and general folder/Windows Explorer shortcuts:

- **F2 : Rename**
- F3 : Find
- ALT+ENTER : Open the properties for the selected object

Note: more keyboard shortcuts are described in this Microsoft article:
http://support.microsoft.com/kb/126449

Windows Basics 101

In order to do anything on a PC, there are some essential skills, **terminology** and **knowledge** that you must master before you'll ever be comfortable.

The things you absolutely **HAVE** to know about **Windows**:

The Desktop

The Windows **Graphical User Interface** (**GUI**) is the **key to getting anything done in Windows**. After you've first logged into your computer, it's divided into two main sections:

- Desktop
- Taskbar

The **Desktop** is the main area of the screen and is basically a storage location. You can store any of the **four types of files there**: **data**, **programs**, **shortcuts** or **folders**. In practice it makes more since to keep your data files somewhere else, but the **Desktop** is a good place for **shortcuts** and **files that you use frequently**.

*Your Desktop is actually a **folder** in Windows.*

NOTE: The desktop is really a **folder** in the **C: drive**. For **Windows XP** that folder is:

- C:\Documents and Settings\username\Desktop

- Create a new text file on your desktop (right-click | New | Text Document)
- Name the file **yourname.txt**
- Open up **My Computer** and navigate to the folder in the above example
- Delete the **yourname.txt** file from that folder
- Go back and look at your desktop

The Taskbar

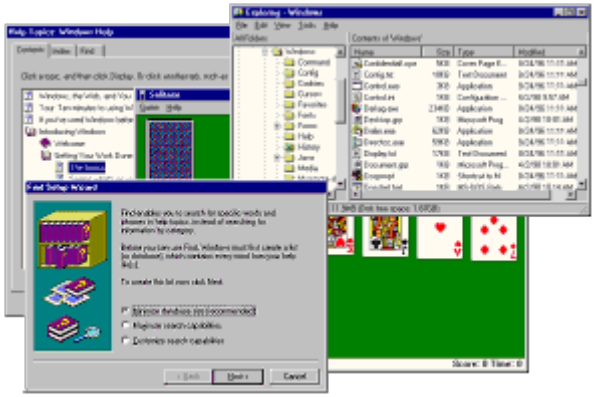
The **Taskbar** is a specialized menu bar in Windows that performs **several functions**. It's the home of the **Start** button (we'll talk more about this later, but it's pretty important so remember where it is). The Taskbar houses the **Quick Launch Menu** which is a storage location for **Program Shortcuts**. You can **move** or **copy** a shortcut from the **Desktop to the Quick Launch Menu** by right clicking and dragging it down there (*remember the right-click-n-drag?*).

Next to the **Quick Launch bar** is an area that holds all of the **open** or **currently running applications** on your computer. A single mouse click (left) on the program name will bring it to the **foreground or pop up the window**. Hitting the little **minus sign** (**minimize**) will **return the program to the Taskbar** or send it to the **background**.

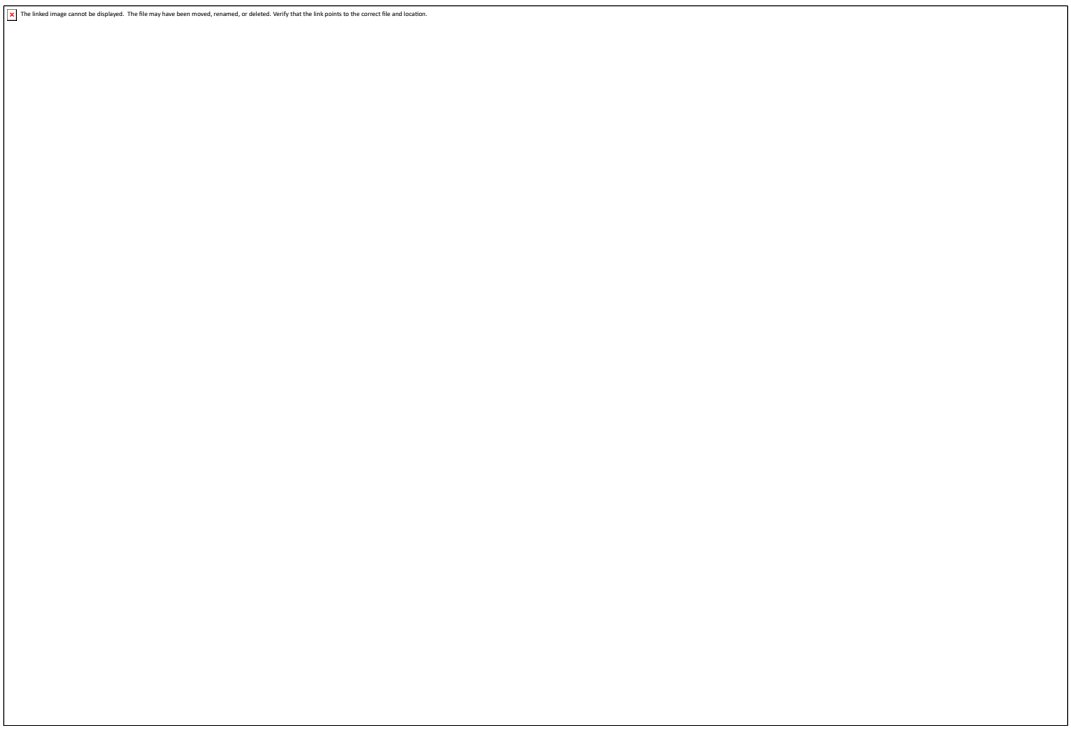
- Right-click on a clear area of the Taskbar.
- Select Properties
- Check (and uncheck) Auto Hide taskbar.

Anatomy of a Window

Windows are what **Windows** is all about. Until you learn how to **comfortably manipulate** Windows; to **open** them, **size** them, **move** them, **close** and **minimize**, you're not going to be very productive.



So let's start out with the **main parts of the Window** in Windows. This will be the **foundation for almost everything else** we do so hang on.



Background & Foreground

When a window is in the **foreground**, it is said to be **active**. It is that is **in front of** any others and the window that the **keyboard and mouse are going to interact with**.

When a window is in the **background**, it is said to be **inactive**. The keyboard and mouse **will not pass any information** to this window. A window can be in the background by simply being **behind another window** or by being **minimized to the Taskbar**.

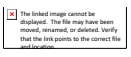
Moving Windows

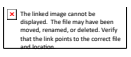
Click and hold on a **clear area of the Title Bar**. Then **drag** the whole window. This **moves** the whole window to a new spot on the desktop. This won't work if the window is in **FULL SCREEN MODE** (see **Manipulating Windows** below)

> ***REMEMBER:*** *Anytime you have **more than one window open** at a time, clicking **anywhere on a window** will bring it to the **front (foreground)** and make it **active**.*

Manipulating Windows

These buttons are your **friends**!





-  This button is the easiest one to understand. Clicking on this button will **close** or **exit** the program or window.
-  This button causes the window to minimize—**go to the background**. This takes it **off of the screen** and puts it down on the **Taskbar**. The program is not closed—it is still running—it's just **not up on the screen** to see.
-  If this button is displayed, it means that the current window is in **FULL SCREEN MODE** and is **taking up the whole screen** and **cannot be resized**. Clicking on this button will change it to the button below and allow you to resize the window.
-  When this button is displayed, the **window is resizable**. Clicking on this button will **toggle to the button above** here and make the window take up the whole screen. (**Double Clicking on the Title Bar** will also toggle to FULL SCREEN MODE)

Resizing Windows

- Verify that the window is **not in FULL SCREEN MODE**: ( is showing)
- Place the cursor over **any corner** until it changes to this:  Click and drag the corner in or out. You can resize in **one or both directions** like this.
- Place the cursor **over the top or bottom edge** until it changes to this:  Then click and drag **up** or **down**.
- Place the cursor **over either side** until it changes to this: Then click and drag **left** or **right**.

- Manipulating & Resizing Windows
- Open up Word, IE and My Computer
- Maximize & minimize each
- Make each non-full screen
- Resize each to about 2/3 the screen
- Manually cascade them over each other then switch between

File Formats

Overview

Whether it's in the **color** arena or a simple **scan to file environment**, we are constantly engaged in conversations with customers about **files**…or their **containers for digital information**.

Some of the deep down technical secrets of files can be pretty daunting, but an overview of the **types**, **terminologies**, **characteristics** as well as the specific file **usages** an **indispensable** part of the **manipulation**, **creation** and **use of digital files**.

Of the **4 types of files** (what are they?) we are going to be focusing on the **data file** in this section.

*The data file is our **container** for **digital information**.*

Graphic File Formats

When we **scan** a piece of paper into a computer or directly to an email, we're converting that piece of paper into a **file**. The **type of file** and thought process surrounding the **selection of that file** are both important factors.

Using the **wrong file type** for a specific application can result in anything from a **slowdown in workflow** to a complete **breakdown** in communication between systems.

For the purpose of this discussion, we're going to concentrate on "**graphic**" file formats - files that are used explicitly for moving around **picture data**. Picture data can represent anything from a **full color photograph** to a **black and white** copy of a **typed** document.

Some common **image file formats** are:

• TIFF	tagged image file format	widely used and flexible
• BMP	bitmapped image	Windows specific
• JPG	Joint Photographic Group	widely used for pictures & web
• GIF	Graphics Interface Format	widely used web file format
• PNG	Portable Network Graphics	emerging web file format
• PICT	Apple Quickdraw metaformat	not as common
• PCX	ZSoft Paint format	older format
• WMF	Windows metafile	Windows raw graphic data

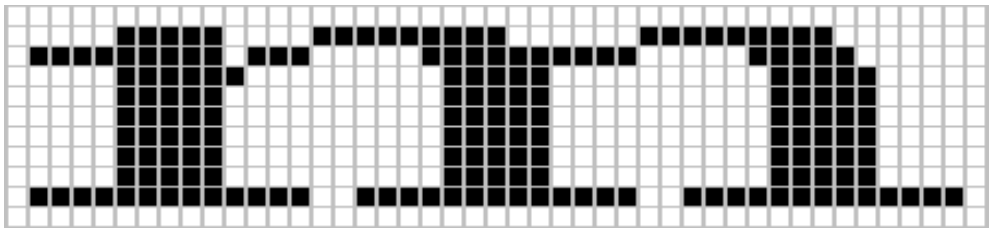
Bit Depth

Different image file formats are capable of holding different amounts of information (**color** or **grayscale** information). Each file format supports a maximum number of "**bits per pixel**."

- 1 bit per pixel refers to an image with 2 colors.
- 4 bits per pixel refers to an image with up to 16 colors or shades of gray.
- 8 bits per pixel refers to an image with up to 256 colors or shades of gray.
- 16 bits per pixel refers to an image with up to 32,768 colors.
- 24 bits per pixel refers to an image with up to 16,777,216 colors.

Raster Image Format

Pixel (PICture ELement) Smallest element of a DIGITIZED image



A **raster image is a set of dots** (or pixels) lined up both horizontally and vertically. A single dot can be 1 bit (on or off, 1 or 0) giving it the capability to represent 2 colors; black & white.

- At 300 dpi (a typical quality for printed documents) an 8½ x 11 document equates to:
- 8.5 x 11 = 93.5 square inches
- 300 x 300 = **90,000 dots** or pixels **per square inch**
- 93.5 x 90,000 = 8,415,000 or **8.4 megabits** (not megabytes)
- 8,415,000/ 8 = 1,051,875 bytes or about **1 megabyte**

An uncompressed, 1 bit, raster image file requires 1M of storage space.

Resolution

The resolution of a file is a measure of **how many dots** are contained in the **original data**. **Not** necessarily how many **dots per inch**, but just **how many** dots.

IMPORTANT – PRINT SIZE VERSES DPI:

Any raster image file has in inherent amount of information contained in the file.

The image file in the above example could be blown up to print **twice its normal size** by **reducing the effective dpi to 150**. The amount of data in the file has not changed. Consequently, **doubling the dpi to 600** would have the opposite effect and the **print size would be half**.

Resolution is frequently stated in **dots per inch** and is a measure of the **image quality** available in the file. Some general **guidelines** apply when thinking about dpi.

- 72 dpi = normal quality onscreen image
- 75 dpi = very low quality print
- 96 dpi = higher resolution computer monitors
- 150 dpi = low quality print
- 200 dpi = fine fax image
- 300 dpi = typical laser print quality
- 600 dpi = high quality laser print
- 1200 dpi = high quality inkjet print (approximates gradations with dots)
- 2400 dpi = extremely high quality; press ready artwork

File Sizes

Here's where the problem starts to come in. A **1 bit, 300 dpi, 8 ½ x 11** scanned image file will be about a **1M file**.

- (8.5 x 300) x (11 x 300) = 8,415,000 bits (1,051,875 bytes)

If that same document is scanned as an **8 bit, grayscale** document then there will be **8 times as much information giving us an 8M file**. The situation is even worse for a **24 bit color** image. Now we've got a **24M file**.

When you talk about storing **a lot of files** or simple **sharing** a file across the **internet** or through **email**, the file size becomes a problem in a hurry. With these numbers, hard drives would **quickly fill up** if something wasn't done to **reduce the size** of the image files.

So what to do?

Compression

File **compression** (or data compression) is a way of **reducing** the size or the amount of a data file.

MP3 is a form of **audio compression** that takes a standard sound wav file (from a CD) and **shrinks it to about 1/10** of its original size. Because of that, **10 times as many** compressed sound files will fit in the same space previously required for the uncompressed files.

There are various types of file compression that we can use to reduce the size of our files. Each type will have positive and negative aspects and the best choice will depend upon the intended use for the documents.

Lossless

The purpose of a compression scheme is to **reduce the size** and consequently the **storage requirements** for our data files. If a compression scheme can be used that shrinks the size of the file for storage and then **returns that file to its original format intact**, this is considered a **lossless** type of compression.

None of the **original information** in the file is **lost** or **removed** in order to shrink the size of the file.

Examples:

- ZIP
- LZW
- Run Length Encoding
- Group 4

Lossy

On the other hand, compression schemes exist that actually **remove data** from the file in order to reduce the size of the file. The **MP3** compression mentioned earlier is a **lossy** compression scheme.

In shrinking the size of the file, information is **lost that can never be retrieved**. Lossy compression schemes are bad for files that need to be opened, **edited** and **resaved frequently**. Each time the file is saved, it is **re-compressed** and **more information is lost** from the file. Almost like making a copy of a copy.

Examples:

- MP3 (Mpeg Layer-3)
- MPG (Mpeg)
- JPEG (Joint Photographic Experts Group)
- JBIG

JPEG compression economizes on the way data is stored and also identifies and **discards extra data**, that is, information beyond what the human eye can see.

Because it **discards** data, the JPEG algorithm is referred to as "**lossy**". This means that once an image has been compressed and then decompressed, it will **not be identical to the original image**.

In general, compressed JPEG images have compression ratios of between **5:1** and **15:1**. A trade-off does exist between the **image quality** and the **amount of compression**.

You don't need to manually **decompress** (or **decode**) images saved in the JPEG format. They are **automatically decompressed** when they are opened with a compatible viewer or editing utility.

Compression Ratios

The compression ratio is a **measure of the compression scheme's ability to shrink a file**. A **2 to 1** (**2:1**) compression ratio means the file will be **half** its original size. **10:1** indicates **1/10 the original size**.

Generally, compression schemes specify a range of ratios because the amount of **compression varies** by **how much information** is on the page.

Compression by File Type

You can actually end up **compressing a file** just by picking a particular **file format** or **setting**. When you select a 1 bit TIFF image, you are reducing the amount of color or grayscale information in the file and therefore applying a type of **lossy compression**.

The original information in the document **will not be contained** with in the scanned image file.

Scanning a **color photographic** into a **grayscale** image is a form of lossy compression- **even though you may not be applying any actual compression scheme**.

Color Concepts

The perception of **color** can be created in different ways. The manner by which the color information starts out has a drastic affect on how it will be **perceived** by our eyes.

RGB

RGB is a term that refers to the creation of color by means of combining the three primary **colors of light**; red, green, blue. RGB color is **created by light**. A computer monitor uses the three colors of light to create roughly 16 million viewable colors. When an image is created by **transmitted light**, you can turn up a single color by simply **projecting more** of that color (or combination) of light.

CMYK

CMYK refers to a process of generating color through the **absorption of light**. CMYK generally refers to the process of **printing** a color document. The **inverse** of the primary colors, Cyan, Magenta and Yellow are combined with Black (k).

A CMYK image (or individual color) can never be any brighter than that of the **surrounding light**. Light is **absorbed** by the printed image and whatever **light is left over** is **reflected** out.

A CMYK image will never be as **bright** and **vibrant** as an **RGB** image because its colors are produced by **reflecting available light** rather than **actively projecting it**.

CMYK files help approximate this **muted**, less-vibrant color, while **still displaying** on an **RGB** monitor.

The Storage Math

Why do we need to know any of this? Well, a **typical scanned** image (**200 dpi, group 4, TIFF**) will take up about **50k** or space (**50,000 bytes**). So when we talk to a customer about **storing electronic files**, it's a good idea to know just a little bit about what that means.

How many **50k** scanned images can we store on a:

- CD 12,800
- 1G 20,000
- DVD 94,000
- 80G 1,600,000

That puts one file drawer (about **2,500 pages**) at about **125M** and **10 four drawer filing cabinets** at about **5G**.

PDF (Portable Document Format)

Adobe Systems is responsible for a **proprietary** file format called **Portable Document Format**. The beauty of this format is that **locks** down the **visual format** of the document. Unlike a webpage that may change based on your **screen size**, **installed fonts** and **browser settings**, **PDF** files are a **snapshot** of the **original** document.

PDF files can **cross platforms** (**MAC**, **PC**, **UNIX**) and because the PDF viewer is **free** from Adobe, the format is extremely **popular** and well **excepted**.

A PDF file may contain **text information** (creating a Word document to PDF) or they may contain **image data** such as a **TIFF image**. PDF files support several different **compression schemes** and because of that, files sizes and print quality can **vary a lot** with this format.

File Usages

So **why** use one file type over another? In a lot of **graphics applications**, this boils down to finding a file format that will **adequately transmit the right information** (color, grayscale, black & white), at the right **resolution**, the right **compression** scheme and in a format that's **compatible** with the **intended use**.

TIFF

This is a popular image format for **scanning** and **document management** applications. All of our MFPs that are capable of **push scanning** to file can create a **TIFF file**.

> *The **Tagged Image File Format** represents a **broad range of standards** and TIFF images created on one system **may or may not be compatible on another**.*

TIFF is a **raster file format** that is generally utilized with one of the lossless compression schemes such as **Group 4** or **LZW**. With Group 4 compression, a typical **200 dpi**, **1 bit** scanned image (uncompressed about 1M) will be **around 50k**. TIFF images are capable of carrying anything from 1 bit B&W images to CMYK color images. TIFF images support both **single** and **multi-page** formats and are the preferred graphic format of many **desktop publishing** applications.

PDF

Portable Document Format files are a good choice when documents are to be **emailed** or **shared with multiple people**, either inside or outside of an organization. Keep in mind that when scanning an image, that scanned image (usually a **TIFF** image) is just **embedded into** the **PDF** file. PDF, in effect, becomes the **wrapper** for the raster image file.

JPG

The **JPG** file format uses the **JPEG compression** scheme which is a **lossy** type of compression. In most cases, this makes for a **poor choice** if the documents are going to be **saved for accuracy** or if the images are ever going to be **edited**.

JPG is a **good choice** for archiving **color photos** because it is optimized for reducing the amount of **color information** in a file. JPEG images can be saved in different compression ratios with the **highest ratios** equating to the **lowest visual quality** (or the most **loss** of data).

GIF

GIF images use a **color pallet** of up to **256 colors**. GIF images are generally thought of as **lossless**, but if you **encode** a full color photograph as a GIF, a **lot of the information will be lost**.

GIF files are good for **web graphics** with **a limited number** of **colors** in them.

Storage & Retrieval

Document Security

Why Document Security?

Even before talking about **document security**, we have to talk about the **why of document security**. Understanding why documents (and/or files) need to be secured is **essential** to being able to make the **right decisions** about security.

- Control **who is allowed to** even **see** a document.
- Ensure the integrity of documents – **don't allow any changes** to the original.
- Control **who has the ability to change** a document.
- Control **who can add** documents to a system.
- Control **who can delete** documents to a system.

Document Level Security

Once a document has been scanned and resides in an electronic file, the information can be **controlled and protected**. Certain individuals may have **complete access** to **add** and **delete** documents while others, more general users, would only have access to view these documents. This could be accomplished with a **kiosk** setup using a single computer for the use by anyone interested in the archives. Certain documents could even be **published to the internet** if desired.

There may be a need to **secure** certain **document level functions** such as:

- Opening
- Modifying
- Printing
- Copying & Pasting

All of these things can be controlled using Acrobat and PDF files (more about this in the labs).

System Level Security

Through **Windows** and the **network**, **permissions** can be set, by **user or group** onto certain **files** and/or **folders**.

- Read Only
- Write or Modify
- Execute

Retrieval

Once the documents are converted to digital format, a suitable **indexing system** will have to be in place to ensure that all documents can be **found quickly** and easily. Selecting an indexing scheme entails the process of **striking a balance** between **implementation time** and **search flexibility**. It may be nice to have 14 different ways to find an individual's records, but the time required to enter that information for every single piece of paper may prove to be cost prohibitive.

In many cases, the same system that was used for the paper documents may also be considered for the electronics versions.

Electronic Document Management Systems

A true **Electronic Document Management (EDM)** software or application program is going to give you the ability to **index** documents. In other words**, index fields** can be applied to documents and then **searched in a database** for retrieving information very quickly.

EDM systems provide added control over the **security** of the documents and may also provide for a **repository** for storing documents or files in a **central location**. This aids in applying security and **backing up documents**.

Indexing Systems

Part of the question about how documents will be used will also entail **how they will be found**. As a collection of documents **grows in size**, it also **grows in the difficulty** to find a particular document.

Indexed Retrieval

Indexing is a process of adding **searchable data**, sometimes referred to as **metadata**, to the scanned image files. This searchable index information is most often stored in a **database** along with the information needed to **quickly retrieve** the needed documents or files.

Indexing can be as simple as **last name, first name** in the case of personnel files or may represent some **unique identifier** like a **social security number** or an **invoice number**.

Business Process Analysis

The choice about **how** and **why to index** archived documents will be determined by the **needs** of the people that will be looking for the documents. A business analyst will work with the customer in order to define a system that is easy to use and works for their needs.

Folder Structure

When you begin **scanning documents** for the first time, it's important to give some thought to the **folder structure** or the storage containers for the documents. The best

starting place for this is the current paper filing system. The folder structure should be laid out in such a way as to provide **easy navigation** to the specific documents needed.

Start with the **largest functional category**; this might be a **physical** storage location like a **records room** or even **a filing cabinet**. The functional area may also be something virtual like **Accounting** or **Payables**.

Think about how the documents will **need to be found**. Think about the **thought process** that you'll go through to **find these documents**. On the surface it may make sense to have separate **Payables**, **Receivables** and **Financial** folders but what's the **first** or **main piece of information** that you'll have? Will you know that it's a financial statement or that it's a financial statement for particular **entity** or **tax id number**?

Indexing & File Naming

Most MFP's today have the ability to **name the file** or scanned image **from the device** itself. But for the most part, it's generally **easier** and **more productive** to rename the files **from the client** or **PC** side.

Files should be named with **sorting** in mind. The first part of a filename will act as **primary sort key** and should reflect how you want the documents to be grouped together.

The **second part of the filename** can then be used for a **secondary key**. This could be something like a **document type** or **description** (financial statement, email, correspondence, etc.)

The file naming convention should be worked out **ahead of time**—prior to beginning scanning—and be well **defined** in the **written procedures**.

We'll talk more about this in the next section: **Electronic Documents Lab**.

The **name** we give a file and the **place we put it** are both important aspects fo the **ease of finding** that document later.

Alpha Sort

Files sorted by name will sort in **alphabetical order**. The underscore symbol will sort before letters.

- _Zen, Bobby.doc
- Doe, John.doc
- Nason, Scott.doc

Primary & Secondary Sort Keys

You can use **multiple pieces of information** as the filename which will contribute to the **ease of finding** and the **ease of identifying** what the file contains. In this example, the

primary search key is the year, then the **document type** then finally **the last name** and **first name**. All **2006 documents** will be **grouped** together.

- 2006_STATEMENTS Nason, Scott.xls
- 2006_STATEMENTS Jackson, Randal.xls

We could also use the **last name as the primary key**:

- Nason, Scott, 2006 STATEMENT.doc
- Nason, Scott, 2007 STATEMENT.doc
- Jackson, Randal, 2006 STATEMENT.doc

In the above example, the documents will be **grouped** together by **last name, first name**.

The Trouble with Numbers

When a PC does an **alpha sort** to display a list of files, **numbers are alphabetized the same way a word would be**. At first, this sounds like a good thing, but it doesn't provide the intended result without a little help.

1, 10 and 100 will all **sort before** 2, 20 and 200. In other words, **anything that starts with a 1 will sort before anything that starts with a 2**. If you want a list of numbered file names to sort in order, you have to **add leading zeros**.

For example: if you know your file list is going to have less than **1,000 files**, then you could **add 2 leading zeros** making **1 i**nto **001** and **10** into **010**.

- file_1.doc
- file_100.doc
- file_102.doc
- file_112.doc
- file_2.doc
- file_20.doc
- file_200.doc
- file_21.doc
- file_3.doc
- file_30.doc
- file_300.doc
- file_31.doc
- file_32.doc
- file_4.doc

Except for Windows XP

The **sorting algorithm** which is used by the operating system to sort file lists was changed in **Windows XP**. A **numeric sort** is used **instead of string sort** - in other words, leading zeros in a file name are ignored.

Don't leave it up to the operating system to sort the files correctly. Always use the proper naming convention to force the sort that you intended.

The Best Naming Method?

So what's the **best method** for **naming** files? The answer is that **it depends on how the documents are to be used**. **How** (and why) will people **be looking for these documents**? Knowing the ways people will look for and **for what purposes**, will aid in the decision.

The most important part of any file naming system is consistency.

Is always a good idea to **write down a specific procedure** for how files are to be **named** including **upper and lower case usage**, **special symbols** and **spaces**.

Electronic Documents Creation Lab

Creating Electronic Documents

We're going to be **using these files in later labs** so make sure and **create** them, **name** them and **save** them **so you can find them again later**.

- Create a **folder** in DDS; "lab-work"
- ----------

Microsoft Word

- Open up **Word**
- Create a new Word document
- Type "APPENDIX" on the top then anything else under that
- Save in **lab-work** folder as **APPENDIX.doc**
- Close Word
- ----------

Windows Notepad

- Open **Notepad** (start | run)
- Create a new Notepad document
- Type "REFERENCE" on the top then anything else under that
- Save in **lab-work** folder as **REFERENCE.txt**
- Close Notepad
- ----------

Microsoft Excel

- Open **Excel**
- Create a new Excel document
- Type "COVER" in the first cell
- Change the font size to really big
- Save in **lab-work** folder as **Cover.xls**
- Close Excel
- ----------

Scanning

Why Scan?

*"Nobody wants to **scan to file**." - Scott Nason*

What are some of the reasons you'd even want to scan? **List some. (remember what a document is)**

What is Scanning?

Scanning is a term that has come to have mainstream acceptance in the office environment today. Often, the term is used interchangeably with **workflow** or office **efficiency**.

Scanning is an extremely **vague** and **generalized** term and by itself, does not represent a specific workflow or increased efficiency. Scanning is nothing more than the:

*"The mechanics of turning **hardcopy** originals into **digital** form."*

Scanning is just a **small piece** of the overall **workflow puzzle**.

Scanning refers to the process of taking **hardcopy** information—usually paper—and turning it into a **digital representation** of that document or original. A scan of a piece of paper is, in effect, a **digital photograph** of that piece of paper. There is no actual "data" contained in the scanned image other than what our brain can make of the **marks on that page**.

The term scanning refers to the **entire process** of taking hardcopy originals, **preparing** them, **scanning** them into a suitable system and providing some **indexing** or **method for retrieving** these documents later.

Archiving

For many years, the preferred method for archiving documents was **microfilm** (or microfiche). Microfilm had the advantage of taking up **less space** than the original paper documents and was generally expected to have a shelf life of anywhere **from 50 to 100 years**. There were several **disadvantages** with microfilm. In order to backup the film, a **photographic copy** process was used that, like copying a copy, **degraded** with every generation. Microfilm could still be **damaged** by the elements and **multiple physical copies** were essential to the integrity of any **disaster recovery** program. Documents **could not be easily shared** and in order to send of a single page to someone, you had to **print** it and either **mail** or **fax** it to the recipient.

With the use of today's technology, paper-based information can be **digitally archived** and saved **indefinitely**. Whether or not a particular medium (i.e. CD, DVD, Optical) will be viable in 50 or 100 years is of little concern anymore. Once the source material is digital, it

can be copied—for **backup & disaster recovery** reasons as well **as information sharing** and **transfer** to newer systems in the future.

Digital documents can be **easily shared** among many users simultaneously, can be **secured** against tampering and unauthorized access and are **easy to back up**.

Informational Scanning

Any archive project begins by determining **why** documents are being saved in the first place. Many documents exist for the sole purpose of saving and/or distributing **information**. These documents have no aesthetic value other than the information they contain. This type of scanning will emphasize the **legibility of the information** contained on these documents and will minimize the appearance of the original document aesthetics (color of paper or print, texture, etc.)

The number one goal of **Informational Archiving** is to **preserve** the integrity of the **content** contained on the physical page.

Aesthetic Scanning

There are documents, on the other hand, that will need to be preserved as close to the original as possible. This includes preserving the **look and feel** of the original document as well as preserving the legibility of the information it contains.

Different techniques and sometimes **different equipment** may be required in order to properly archive these originals. More care and **more technical experience** are required on the part of the archivist to accurately save these documents. The individuals performing this type of archiving need to be **thoroughly knowledgeable** in the various technical aspects of scanning file formats and digital file compression.

Scanner Basics

Basic Scanner Specifications or Features

- PPM/ IPM
- DPI
- Types of Paper Feed
- Paper Sizes & Capacities
- Interface
- Driver
- Special Features

PPM

Scanners are designed for turning physical media—usually paper—into an electronic file. Scanners are most commonly rated in **PPM** or **pages per minute**. You may also see the term **IPM** or **images per minute**. A single **duplex** page **contains two images**; the front side and the back side.

DPI

A scanner's resolution will be rated in **DPI** or **dots per inch specification**. The DPI is a **rough indicator of the quality** of a scanner (but it's not the only indicator). Watch out for **optical** (or native) resolution verses **interpolated** resolution. Some scanners may advertise **2400 DPI** but only have 600 DPI optical. The **real DPI** is 600 while the 2400 is accomplished through **software enhancements** that give the impression of 2400 DPI.

Types of Feeders

For the most part, scanners are available with the same **feeding options** as **fax machines and high speed copiers**. Automatic feeders are available that can handle as many as five hundred sheets.

And like copiers, scanners are available with **flat beds** or platen (glass) tops to allow for **odd shape originals or bulky objects** such as books to be scanned.

The type of feeder you chose will be related to the types of documents the scanner will be required to handle.

Paper Sizes & Capacities

What are the **minimum** and **maximum paper sizes** and **weights** that can be fed through this scanner? Is it different for the **auto feed** verses manually **on the glass** (if those options exist)? **How many sheets** of what **weight** paper will the auto feeder hold?

Interface

A scanner's interface determines how it **physically connects** to either a PC or network. The traditional standard for scanners was the **SCSI** (pronounced scuzee) interface. This is an actual card that has to be installed in your computer. SCSI devices connect with a special type of SCSI cable and its length is limited to about six feet. SCSI scanners are a little bit **harder to configure and install** than USB devices but their **performance and stability** are excellent.

More commonly, scanners are now coming standard with **USB** interfaces that allow for easy connection to most computers. Keep in mind that there's a **USB 1.1** and a **USB 2.0**. USB **1.1 is the older**, **slower** standard and its speed is not suitable for any production desktop scanning application. **Only use USB 2.0** for high speed production scanning.

Driver

TWAIN or **ISIS** are by far the most common drivers that will come with a scanner. The driver will determine **what software programs will work with this device**. We'll talk **more** about this **later**.

Special Functions

Besides handling paper, there are specialty features that some scanners offer to enhance the scanning process.

- Wide Format (11 x 17 and larger)
- Imprinting
- Bates Stamping
- Endorsing

Scanning Quality

Image Quality

Because in scanning, we are actually **converting** from one medium to another, **there will always be** some **loss of quality**.

> *The process of **quality control** (among other things) is really the process of **minimizing this loss of quality**.*

This starts by understanding the **key concepts that affect the quality** of the scanned image and how they relate to each other.

DPI

The resolution or DPI (dots per inch) along with the bit depth of a scan will determine its **maximum available quality** and have an effect on the eventual **file size** of the digital archive. (*list from earlier discussion*)

- 72 dpi = normal quality onscreen image
- 75 dpi = very low quality print
- 96 dpi = higher resolution computer monitors
- 150 dpi = low quality print
- 200 dpi = fine fax image
- 300 dpi = typical laser print quality
- 600 dpi = high quality laser print
- 1200 dpi = high quality inkjet print (approximates gradations with dots)
- 2400 dpi = extremely high quality; press ready artwork

Compression Type

Make sure and choose a **file** and **compression** type that is **appropriate to the type of original** and the **reason** it's being scanned. **Text documents** could use a **1 bit, black & white TIFF image can use an LZW lossless compression**. **Color** images would benefit from **JPG** (lossy) compression but the **percentage** setting will affect the quality.

PULL verses PUSH Scanning

All scanning utilizing, either a TWAIN or ISIS driver, initiates as a **PULL SCAN**. A host application must be opened and input "pulled in" from a scanning device. Issues surrounding Pull Scanning are generally less technical and more "**workflow**" related. Pull Scanning is a way of performing a task that causes someone to interact physically with two different devices; the scanner (or MFP) and the host application (or computer).

A **PUSH SCAN**, on the other hand, allows the user to complete the entire scanning operation *from one location*. This eliminates, not only a step, but the need for the physical movement between devices.

TWAIN verses ISIS

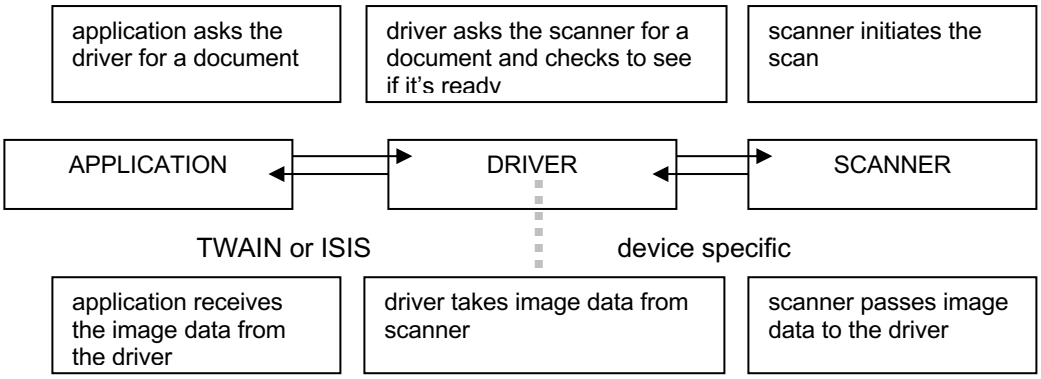
Scanner Drivers

Scanners require **drivers**, just like printers do. A scanner driver "talks" to the scanner and **interprets** for the operating system (Windows) or for a **particular program** (like Adobe Acrobat).

Scanner drivers allow for accessing and setting all of the unique settings and features available to a particular scanner. From **DPI**, to **document size** to special features like **deskew** and **despeckle**. All of these are set in the particular scanner driver that comes with the scanner.

Scanner drivers come in two main "flavors"; **TWAIN** and **ISIS**. Basically you can think of these as a **language**—like **French** or **German**. The only thing you really need to know is whether or not **your application program or software is capable of speaking this language**. There are some programs (like Adobe Acrobat) that support only scanners that have TWAIN drivers. There are other programs like **Canon's imageWARE** that support **both TWAIN and ISIS** scanners so you have more to chose from when picking a scanner to use with that system.

A scanner **driver** is simply an **interface** between an application or operating system and the device itself. The driver **knows how to "talk to" the device** and how to talk to the host application. This allows one application to support an unlimited number of scanners by simply supporting the scanner driver specification (TWAIN or ISIS).



TWAIN and **ISIS** drivers are **not compatible** with each other.

In general, **TWAIN** drivers have traditionally been associated with **the lower end, consumer** devices, while **ISIS** has been found more in the **high-end, high speed** applications and devices.

Digital Workflow

Digital Workflow is a rather generic term that refers (loosely) to the **process of moving an organization's information around electronically**. For our discussion, this focuses on the process for dealing with electronic documents.

The electronic document workflow will generally be **a set of procedures** that govern things like **file naming** conventions, **storage locations** on the network, **backup** responsibilities and so on. Certain workflows may specify the use of **barcodes** for indexing or tracking documents.

Document Preparation

Document prep refers to the process of **getting the documents ready for scanning**. Document prep can involve **removing files** from a folder, **pulling staples** and **straightening** curled edges.

Document prep may entail adding **cover sheets**, separating **simplex** and **duplex** documents or dividing pages by **document size**.

On **larger jobs**, there may be **separate individuals** assigned to document preparation and document scanning.

Driver (Pull Scan) Lab

- Using the Driver (TWAIN pull scanning)
- Go through settings – scan at various dpi, feeder selection, on-off-line, etc.
- Grayscale, black & white, color (if available)

Push Scanning Lab

- Send a document to your own desktop/ folder.
- Send a document to someone else.
- Send more than one page.
- Explore the send options (per device basis)

Adobe Acrobat Lab

Overview

Adobe Acrobat provides a **platform independent** method of exchanging documents between various **computers** and **users**. Documents may be **secured** to various degrees and you can **lock down the format** of the original document so that the pages are presented **the same on different computer monitors and printers**.

Main Acrobat Advantages:

- **locks down** the document **format** (pagination or page breaks)
- can embed **fonts** (a setting)
- can embed **graphics** (quality is settable)
- Reader is readily **available** (easily understood and distributed free)
- advanced **security** features
- easy **page level** manipulation (**Acrobat**)

Brief History of the Internet

- 1957
  - The USSR launches **Sputnik**, the first artificial earth satellite. In response, the United States forms the **Advanced Research Projects Agency** (ARPA) within the Department of Defense (DOD) to establish US lead in science and technology applicable to the military.
- 1962
  - RAND Paul Baran, of the RAND Corporation (a government agency), was commissioned by the U.S. Air Force to do a study on how it could maintain **its command and control** over its missiles and bombers, after a nuclear attack. This was to be **a military research network** that could survive a nuclear strike, **decentralized** so that if any locations (cities) in the U.S. were attacked, the military could still have control of nuclear arms for a counter-attack.
- 1968
  - ARPA awarded the **ARPANET** contract to BBN. BBN had selected a Honeywell minicomputer as the base on which they would build the switch. The physical network was constructed in **1969**, linking four nodes: University of California at Los Angeles, SRI (in Stanford), University of California at Santa Barbara, and University of Utah. The network was wired together via 50 Kbps circuits.
  - Backbone: **50Kbps** ARPANET **- Hosts: 4**
- 1992
  - **World-Wide Web** released by CERN.
- 1994
  - Pizza Hut offers pizza ordering on its Web page.

What is HTML (Hypertext Markup Language)

First internet documents were plain *text* only. Minimal formatting (**bold**, underline, *italics*). HTML primary design was for structural markup not visual presentation.

That means that web pages looked different based on your system settings, screen size & resolution, etc.

PROBLEMS: Individual users **fonts** were different, different brands of **browsers** display the page differently, screen & browser **sizing** will reformat the text – limited **placement** options.

What was missing was a way to **EXCHANGE DOCUMENTS** THAT **LOOKED THE SAME** AS THE **ORIGINAL**.

The problem **still exists today** with documents like **Microsoft Word** and **Quark Xpress**. Word will **substitute fonts** if they're not on the system that's viewing the document and it **won't even give you an error message**.

Quark Xpress keeps **fonts** and **graphic** files **out of the actual document**. The document itself only contains **pointers to the image and font files**. Quark will at least **warn you** that fonts and or graphic files are missing. But you may still be left with an **incomplete document**.

Portable Document Format

Adobe Systems has pioneered the idea of the "**portable document format**" or **PDF** file. PDF first appeared in **1992** and Adobe took the lead early on by providing a means for transporting **platform independent** documents with the formatting (**fonts**, **margins**, **page breaks**, **graphics**) intact. Through the use of a **free** portable document viewer (**Acrobat Reader**), Adobe has successfully positioned PDF as the accepted standard for **document portability**.

The **full version of Acrobat** has powerful features for document **manipulation**, **management**, **storage** and **retrieval**. Acrobat is **affordable** and a **highly effective** first step toward more robust document management.

Acrobat Quality

The Debate

QUESTION: "Are Acrobat documents capable of providing high enough quality to be used for anything other than web documents?"

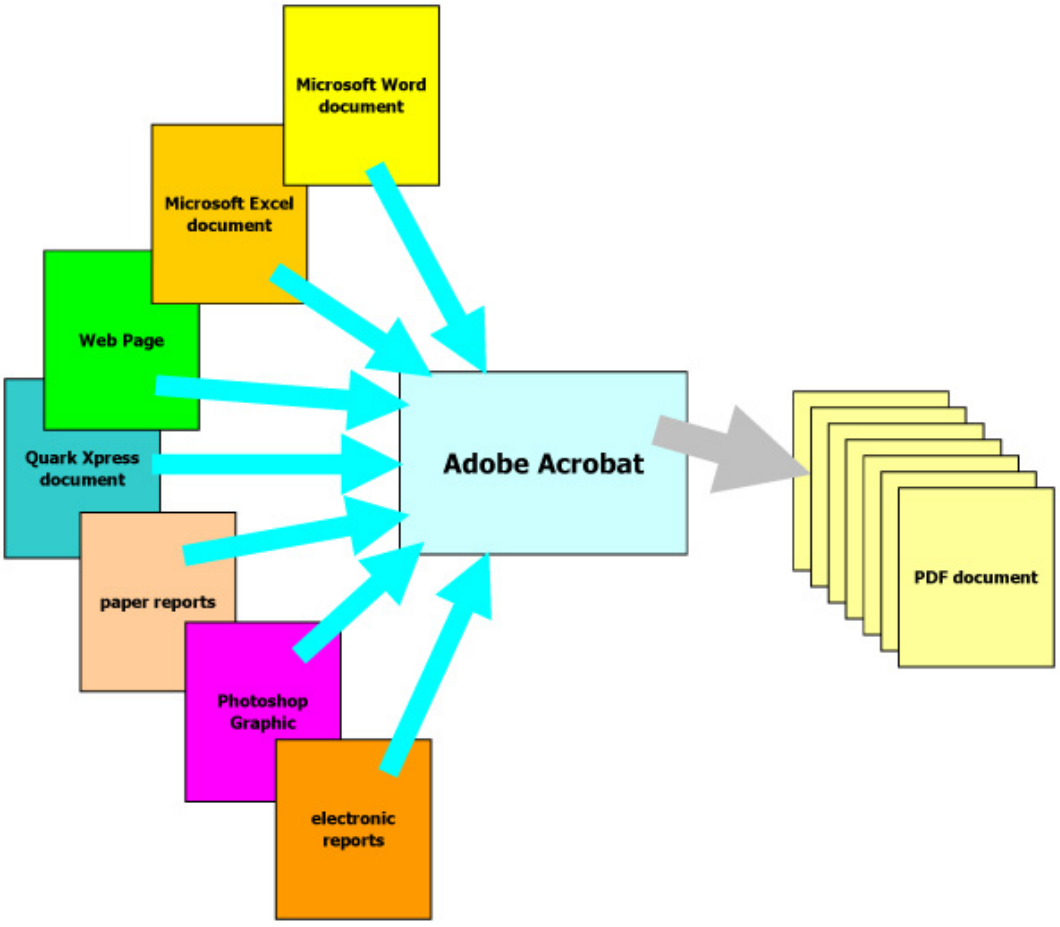
ANSWER: "Yes."

The debate over Acrobat quality is caused by **confusion** and **lack of knowledge**. The first widespread use of PDF files were for internet delivery – which is **optimized for download speed**. This has an effect on the quality of these documents.

Adobe Acrobat has the **ability** to create documents that can be used **for high quality printed and press** work. They can provide **1200 dpi** of images and graphics and even **embed advanced color information** for color patching on a press or other output devices.

Acrobat Workflow

What does that mean – simple explanation (Acrobat at the center – combine **electronic** and **paper**)



PDF files become the **transport medium** behind multiple types of shared **information**.

- Electronics files like Word & Excel
- Web pages
- Database Reports
- Accounting; Inventory and other Reports
- Scanned Image Files

Electronic verses Image Files

There are really **two** main types of **PDF** files:

- Those created from **Electronic** Documents
- Those created from **Scanned** Images

Electronic Document File PDFs

These files **contain the original electronic text** from the original.

PDF files created from **Electronic Documents** are different from **scanned image files** in that they **contain editable text**. If the **security** of the Acrobat file **allows it**, these files can be edited (a little). This **text can be searched** to find a particular word or phrase.

Scanned Image File PDFs

PDF files created from **Scanned Image** files are like **digital photos**. They're nothing more than a **picture** or representation of **what was on the paper**. To the computer, these are **not text documents**; they **do not contain any editable or searchable text**.

The computer doesn't see the image data as information; only we do.

Searchable PDF (OCR)

Acrobat Professional can run **OCR** (**O**ptical **C**haracter **R**ecognition) on scanned files, creating **searchable** documents **from scanned image** files. Through the use of **third party plug-ins** and more **advanced OCR engines**, Acrobat's internal capabilities can be easily **enhanced** or extended.

Remember, PDF files that are electronically created (i.e. by printing from something like Microsoft Word) will always be searchable. It's only the scanned image files that need to be OCR'd.

Acrobat Family

Reader verses **Professional** (or **Standard**)

- Acrobat **Reader**
- Acrobat Elements
- **Acrobat Standard**
- **Acrobat Professional**
- Acrobat 3D

Brief outline of features – what's covered here what's not.

Electronic PDF Creation Lab

Electronic PDF from Microsoft Word

- Open the APPENDIX.doc file from the earlier exercise
- Print to PDF (Take a simplified look at the Printer Settings available)
- Save as APPENDIX.pdf

Electronic PDF from Microsoft Word (single pages)

- Open the Word doc; originals-Word.doc from the CD "lab-documents" folder
- Print each page to a separate file (Word-2.pdf. Word-5.pdf, Word-9.pdf)

Electronic PDF from Windows Notepad

- Open the REFERENCE.txt from the earlier exercise (Start | Run | Notepad)
- Print to PDF
- Save As REFERENCE.pdf

Electronic PDF from Microsoft Excel

- Open the Excel Cover.xls file from the earlier exercise
- Print to PDF
- Save As Cover.pdf

Electronic PDF from Internet Explorer

- Open Internet Explorer
- Print to PDF
- Save As Web Page.pdf

Pull Scanning into Acrobat

Pull Scan Practice

- (TWAIN PULL – user either standalone scanner of MFP)
- If necessary, print classroom originals from originals-scanning.pdf.
- Scan single page
- Scan multiple pages
- Scan double sided pages
- Scan Black & White; Grayscale

Scan the Master Working Document

- Scan classroom originals (complete set)
- Save As master_document.pdf in the lab-work folder.

Combining Digital Documents

The **power** of an **Acrobat Workflow** lies in the ability to **manipulate** a document at the **page level**. This means that we can do things to the document such as moving and deleting pages.

We're **not going to use Acrobat to perform any editing on individual pages**—that will be left for the program the files came from (i.e. **Microsoft Word**).

Types of Page Level Manipulation:

- Move pages
- Insert pages (from file)
- Insert pages (from scan)
- Delete pages
- Extract pages

Moving Pages

- Move page 3 to front
- Move page 1 to end

Inserting Pages

- Insert Word-2.pdf and Word-5.pdf after pages 2 and 5.
- Insert Word-9.pdf at the end of the document
- Insert Excel cover in beginning
- Insert Notepad Reference page
- Insert Word Appendix page

Deleting Pages

- Delete page 2 from the original scan – PULL SCANNING INTO ACROBAT.

Extracting Pages

- Extract page 5 from the original scan – PULL SCANNING INTO ACROBAT. Select DELETE to remove the page from this document after it's extracted.
- Save as: LabPage5.pdf.

Advanced Features (optional)

Briefly – (*not intended to be complete lesson, just an overview*)

Through the use of **Acrobat's advanced features** such as **bookmarks**, **commenting**, **cataloging** and **indexing**, paper-based documentation can truly begin the journey toward **electronic knowledge**.

Acrobat's ease of use, advanced feature set and almost **universal acceptance**, all combine to make this an extremely powerful tool.

- Rotating pages.
- Create 4 bookmarks
- Move a bookmark under another bookmark
- Add a comment
- Create a searchable PDF by OCRing
- Document Level Security

Summary – QA

- Digital Documents are created and used every day.
- A document is a container for information.
- You have to understand Windows in order to deal effectively with digital documents.
- Acrobat can be at the center of a powerful digital workflow.

QUESTIONS

Appendix A – Classroom CD Contents

Classroom CD folder structure:

- Handouts
  - Sniglette #1) File System
  - Sniglette #2) Device Paths
  - Sniglette #3) Computer Math
  - Sniglette #4) Scan to File
  - Glossary
- Lab-documents
  - FOLDER: file naming
  - FOLDER: mouse selection exercises
  - Originals-scanning.pdf
  - Originals-Word.doc
- Manual
  - This manual; **dds-level1_manual.pdf.**
- Resources
  - Acrobat Product Comparison (acrobat8_matrix.pdf)
  - File Format Reference.doc
  - The history of PDF.doc

Appendix B – Acrobat Product Comparison

Acrobat 8 features	Reader	Elements	Standard	Pro	3D
View, print, and search Adobe PDF files	yes	yes	yes	yes	yes
Use Start Meeting to collaborate online and share documents in real time with Adobe Acrobat Connect™ software	yes	yes	yes	yes	yes
Create PDF documents from any application that prints, including one-button creation from Microsoft Word, Excel, and PowerPoint	no	yes	yes	yes	yes
Protect PDF documents with passwords and 128-bit encryption	no	yes	yes	yes	yes
Apply restrictions on printing, copying, and altering PDF documents	no	yes	yes	yes	yes
Create PDF documents with one-button ease from Microsoft Outlook, Internet Explorer, Publisher, and Access, as well as Lotus Notes	no	no	yes	yes	yes
Combine files from multiple applications into a single PDF document	no	no	yes	yes	yes
Combine multiple files into a searchable, sortable PDF package that maintains the individual security settings and digital signatures of the original PDF files	no	no	yes	yes	yes
Use familiar commenting tools including sticky notes, highlighter, lines, shapes, and stamps	no	no	yes	yes	yes
Conduct shared document reviews that allow review participants to see one another's comments	no	no	yes	yes	yes
Compile comments from all reviewers into a single PDF document and sort, filter, and print as needed for easy reconciliation	no	no	yes	yes	yes
Authenticate and certify PDF documents with digital signatures	no	no	yes	yes	yes
Scan paper documents into PDF and automatically recognize text with optical character recognition (OCR)	no	no	yes	yes	yes
Save Adobe PDF files as Microsoft Word documents, retaining the layout, fonts, formatting, and tables, to facilitate reuse of content	no	no	yes	yes	yes
Complete tasks more quickly with a streamlined user interface, new customizable toolbars, and a "Getting Started" page to visually direct you to commonly used features	no	no	yes	yes	yes
Find hidden information, including metadata, annotations, attachments, form fields, layers, and bookmarks, and delete as needed	no	no	yes	yes	yes
Permanently delete sensitive information, including specific text or illustrations, with redaction tools	no	no	no	yes	yes
Create PDF documents with one-button ease from AutoCAD®, Microsoft Visio, and Microsoft Project (Windows® only)	no	no	no	yes	yes
Preserve document layers in technical drawings in Visio and AutoCAD, and object data in Visio (Windows only)	no	no	no	yes	yes
Create fillable PDF forms from scanned paper, existing PDF documents, Microsoft Word documents, or Excel spreadsheets	no	no	no	yes	yes
Automatically recognize form fields on static PDF documents and convert them to interactive fields that can be filled electronically by anyone using free Adobe Reader® software†	no	no	no	yes	yes

Enable Adobe Reader† users to participate in reviews with complete commenting and markup tools, including sticky notes, highlighter, lines, shapes, and stamps	no	no	no	yes	yes
Enable Adobe Reader† users to fill and save PDF forms locally for offline use‡	no	no	no	yes	yes
Enable Adobe Reader† users to digitally sign PDF documents	no	no	no	yes	yes
Convert 3D designs from major CAD applications to PDF documents for use by extended teams without expensive CAD or viewer software	no	no	no	no	yes
Insert 3D CAD designs into Microsoft Word, Excel, and PowerPoint documents and convert the files to PDF	no	no	no	no	yes
Create rich, interactive documents with 3D content: Add materials, create exploded views, edit lighting, and save as 3D objects or 2D raster/vector images for use by extended teams	no	no	no	no	yes
Include PMI data with 3D designs in PDF documents for use by extended teams	no	no	no	no	yes
Export precise manufacturing data from PDF files into standard 3D formats such as STEP, IGES, and Parasolid for use in CAM and CAE applications	no				